A Lexical Theory of Phrasal Idioms

Paul Kay

U.C. Berkeley & Stanford University

Ivan A. Sag

Stanford University

Abstract

This paper presents a theory of syntactically flexible phrasal idioms that explains the properties of such phrases, e.g. *keep tabs on*, *spill the beans* in terms of general combinatoric restrictions on the individual idiomatic words (more precisely, the *lexeme*s) that they contain, e.g. *keep*, *tabs*, *on*. Our lexical approach, taken together with a construction-based grammar, provides a natural account of the different classes of English idioms and of the idiosyncrasies distinguishing among particular phrasal idioms.

1 Introduction

Among the most important desiderata for a theory of idioms (also known as multi-word expressions (MWEs) are the following four. First, idioms divide into classes whose distinct properties (see below) need to be theoretically accommodated. Second, the theory should get the semantics right. It should, for example, represent the fact that (1a) and (1b) have roughly the same meaning, which can be glossed as (1c):[1]

(1) a. let the cat out of the bag

---

[1]All unstarred examples in the remainder of this paper are attested on the web, unless otherwise noted.

b. spill the beans

c. reveal a/the secret(s)

Third, the theory should get the syntax right. For example, it should predict that (2a) is utterly unacceptable while (2b) is not, despite the fact that (3a) and (3b) mean about the same thing.[2]

(2) a.*The bucket was kicked . . .

b. The ghost was given up . . .

(3) a. kick the bucket

b. give up the ghost

Finally, the theory of idioms should be part and parcel of a general theory of grammar; the theory of idioms should add as little machinery to one's overall grammatical approach as possible, ideally nothing. We will expand on all these desiderata using the framework of Sign-Based Construction Grammar (SBCG), as developed in Boas and Sag 2012 (see in particular Sag 2012). Our theory of idioms will be lexicalist; in particular we will divide the lexicon into words that appear only in idioms, such as *cat, bag, spill* and *beans* in (1) and the words that are pronounced exactly like these that appear in nonidiom contexts like (4) (see Nunberg et al. 1994 [henceforth NSW] and also Wasow et al. 1984):

(4) The vet held onto the dog and let the cat out of the bag. (invented example)

---

[2] It is difficult to find two idioms with similar extensions only one of which passivizes. For readers who find (2b) unacceptable, a quick Google search will reveal that examples of the passive form of this idiom are not hard to find. Most examples appear to be metaphorical, about abandonment of projects or activities rather than actual deaths, but there are some literal examples as well, e.g.

(i) Then, in January of '72, the ghost was given up. It's still difficult to write about Patchen's death.

(ii) Gazing into that jar . . . the poison as pure as Alpine snow . . . he had witnessed . . . the awful instant when the ghost was given up, the wings settling down into the first magnificence of death.

We will in fact propose a three-way split in the lexicon. In addition to nonidiomatic words and the words occurring only as arguments of idiom predicators, there are the idiom predicators themselves, such as *let* and *spill* in (1). Phrases headed by these words are not restricted to occurrence in idiom contexts (*Kim spilled the beans* can be embedded in a nonidiom context, as in *Someone said Kim spilled the beans.*). The words like idiom *spill* pattern with nonidioms in this regard. But these words play the special role in phrasal idioms of requiring some or all of their argument phrases to be headed by idiom words. An idiom predicator thus bridges between the idiom context it governs and the nonidiom context in which the phrases that it projects occur.

A key problem in formulating a theory of idioms which, like ours, accords meaning to many idiom words and concomitantly allows a certain amount of syntactic freedom to the occurrences of these words, is to limit that freedom of occurrence appropriately. For example if we analyze *beans* of *spill the beans* as meaningful (meaning something like 'secret(s)') and idiomatic *spill* as meaning something like 'reveal', and use these assumptions to explain why idiomatic *beans* can be modified and *spill* can be passivized, as in (5), we must also prevent *beans* meaning 'secret(s)' from occurring in environments like (6), where *beans* is not lexically licensed by idiomatic *spill*:

(5) The Pam Anderson beans were spilled long ago, but who's joining the Baywatch babe in the Dancing With the Stars spotlight this year?

(6) *John Dean refused to keep the Watergate beans.

Several dimensions of difference among idioms have been observed in the literature (For reviews, see Fillmore et al. 1988 and Sag et al. 2002 and the references cited therein). The dimension we will focus on primarily is the general correlation between the syntactic plasticity of an idiom and its semantic compositionality (noted originally by NSW). In footnote 2 we mentioned that *kick* of *kick the bucket* does not passivize, whereas *give* of *give up the ghost* shows some signs of passivizing. We noted in the text that these expressions convey roughly the same meaning but, crucially, in the *kick the bucket* case the meaning is presumably a simple predicate 'die', while in the *give up the ghost* case it appears to be a more complex, decomposable predicate like 'lose (one's) life/soul'. NSW's hypothesis, which we pursue further here, is that meaningfulness of the words of an idiom correlates with their syntactic potential; specifically, meaningful idiom words can be modified and displaced (more precisely, can appear in more than one syntactic context), whereas meaningless ones cannot.

Following Bauer (1983) and Sag et al. (2002), it is useful to distinguish at least three types of (English) idioms:

Fixed Expressions: Expressions which appear to contain more than one word but which display idiosyncratic syntactic combination (and *a fortiori* no semantic compositionality). Examples include *by and large, right away, first off, all of a sudden*. These can, without loss of generality, be listed as single words in the lexicon, despite their spelling suggesting a multiword history.

Semi-Fixed Expressions: Idioms which obey normal syntactic constraints but which are nonetheless quite frozen as well as semantically non-compositional. Examples are *black and white, kith and kin, bright-eyed and bushy-tailed, it takes one to know one*. One is tempted to accord members of this class the syntactic structures that they appear to exemplify, but since they strongly resist both morphological syntactic manipulation, parsimony argues for encoding them as single words. However, since they can be interrupted by epistemic or intensifying adverbs, they cannot be considered single words:[3]

(7) a. It sure/surely/certainly takes one to know one.

    b. It definitely takes one to know one.

    c. bright eyed and totally/completely bushy tailed

The class of semi-fixed expressions also includes idioms that are syntactically frozen and semantically noncompositional but morphologically alternating. Relevant examples are **kick/ kicks/kicked/kicking** *the bucket* and **trip/tripped/tripping** *the light fantastic*. The inflectional potential of *kick*, for example, shows that *kicked the bucket* is not a fixed expression, but rather a verb phrase constructed according to the familiar English pattern. After introducing the next and final type, we return to the analysis of this and other semi-fixed expressions.

Syntactically Flexible Expressions: Idioms with the following properties:

(8) a. they exhibit modulations of syntax or morphology (or both),

---

[3]The appearance of 'epithets' within these idioms, as in *kith and* **bloody** *kin*, cannot be taken as evidence of phrasal structure because, as is well known, epithets are also allowed within uncontroversial words, e.g. *out-***bloody***-rageous*, *fan-***freakin***-tastic*.

b. their meaning cannot be composed from the 'literal' (nonidiomatic) interpretations of their parts, but

c. they are analyzable in terms of parts whose special (idiomatic) meanings can be modified or quantified under certain circumstances.

Parade examples are *pull strings* and *spill the beans*, which have been analyzed as containing special idiom words *pull* 'manipulate', *strings* 'access', and *spill* 'reveal', *beans* 'secret(s)'. Idioms of this type are widely taken to have the same syntactic structure as the homophonous nonidiomatic expressions, which in turn explains why they can be inflected, modified and dislocated syntactically. This is the type of idiom that we will be mostly concerned with in this paper.

2   The Analysis of Syntactically Flexible Expressions

We begin with some basics of SBCG, which is equipped with a multiple-inheritance hierarchy of typed features structures of the familiar kind.[4] Readers familiar with HPSG or with the kind of Berkeley Construction Grammar presented in, for example, Kay and Fillmore 1999, should find nothing very unfamiliar in the tools and notations introduced in this paper.[5]

The most important type of feature structure in SBCG is *sign*, with its various subtypes *word*, *lexeme* and *phrase*. Each sign has a FORM feature whose value is a morphological representation of the expression, notated here in standard English orthography. Two other features of the sign are SYNtax and SEMantics. The value of SYN is a feature structure that specifies a value for features like CATegory and VALence; we return to SEM below. The CAT values are feature structures (assigned to various types of category) that specify values for appropriate features, including Lexical Identity (LID), whose value is a list of frames (typically a singleton list). CAT specifications 'percolate' in headed constructs from the head lexeme to higher projections (See Sag 2012).

The VAL feature takes as value a list of signs, corresponding to the elements that an expression can combine with (in order of decreasing obliqueness, or accessibility) the syntactico-semantic arguments of each predicator (⟨Subject, Direct Object, ...⟩). Expressions like NPs and clauses have the empty list as their VAL

---

[4]See Sag et al. 2003 for an elementary presentation.

[5]For those familiar with the latter but not the former, read the embedded bracketed attribute-value matrices (AVMs) as embedded boxes. Angled brackets (⟨...⟩) denote lists.

value, as they are intuitively 'saturated', i.e. they already contain (canonically, at least) all the predicator's arguments.

The values of SEM include specifications for the features INDEX and FRAMES. We assume an indefinitely large list of referential indices 1,2,.... Non-referring expressions such as idiomatic *bucket* receive the specification [SEM [INDEX *none*]]. The FRAMES feature takes a list of elementary predications (roughly the RELS of Minimal Recursion Semantics MRS[6] as its value. In certain cases, the FRAMES list of an expression is empty.

Our analysis appeals to a frame-based conception of semantics (Fillmore 1982, 1985, Fillmore and Baker 2010), in which we assume a bifurcation of the universe of frames into canonical frames (*c-frame*s) and idiom-frames (*i-frames*). For example, *i-beans[secret]-fr* models the idiomatic sense of *beans* in *spill the beans*. It is an *i-frame* that can be loosely paraphrased as *secret* and is distinct from the nonidiomatic sense of *beans*, which we represent simply as *beans-fr*, which is a *c-frame*.[7] Lexemes are identified in our theory in terms of a feature LEXICAL-ID (LID), whose value identifies each lexeme (and the phrases it projects) by its principal frame.[8] In the case of nonidiomatic lexemes and meaningful idiomatic lexemes, the LID value is the main frame of the lexeme.[9]

(9)

$$
\begin{bmatrix}
word \\
\text{FORM} \quad \langle\text{beans}\rangle \\
\text{SYN NP}\begin{bmatrix}\text{LID} & \left\langle \boxed{1}\begin{bmatrix}beans\text{-}fr \\ \text{THEME} & x\end{bmatrix}\right\rangle\end{bmatrix} \\
\text{SEM} \quad \begin{bmatrix}\text{INDEX} & x \\ \text{FRAMES} & \langle\boxed{1}\rangle\end{bmatrix}
\end{bmatrix}
\qquad
\begin{bmatrix}
word \\
\text{FORM} \quad \langle\text{beans}\rangle \\
\text{SYN NP}\begin{bmatrix}\text{LID} & \left\langle \boxed{1}\begin{bmatrix}i\text{-}beans[secret]\text{-}fr \\ \text{THEME} & x\end{bmatrix}\right\rangle\end{bmatrix} \\
\text{SEM} \quad \begin{bmatrix}\text{INDEX} & x \\ \text{FRAMES} & \langle\boxed{1}\rangle\end{bmatrix}
\end{bmatrix}
$$

---

[6]See Copestake et al. 2005.

[7]Although the idiom word *beans* appears to have plural form, the corresponding *i-beans-fr* 'secret or secrets' denotes an abstract (mass) entity.

[8]For readers familiar with HPSG, the LID feature subsumes the function of the PFORM feature and accomplishes the analogous function for lexical heads of all categories.

[9]For conciseness, we follow the abbreviation conventions of Sag 2012. Thus, (i), depending on the context, abbreviates either (ii) or (iii) and (iv) abbreviates (v):

(i)   NP[LID L]   (ii)
$$
\begin{bmatrix}
sign \\
\text{SYN}\begin{bmatrix}\text{CAT}\begin{bmatrix}noun \\ \text{LID} & L\end{bmatrix} \\ \text{VALENCE} \langle\rangle\end{bmatrix}
\end{bmatrix}
$$
(iii)
$$
\begin{bmatrix}
syntax\text{-}object \\
\text{CAT}\begin{bmatrix}noun \\ \text{LID} & L\end{bmatrix} \\
\text{VALENCE} \langle\rangle
\end{bmatrix}
$$

(iv) NP_i (v)   NP & [SEM [INDEX i]]

The phrases these words project are distinguished as nonidiomatic or idiomatic in terms of their LID value, because the LID value of the lexical head is identified with that of the phrase it heads by the Head Feature Principle (Pollard and Sag 1994) or some similar constraint, as indicated in (10):

(10)

$$
\begin{bmatrix}
\text{FORM} & \langle \text{the, beans} \rangle \\
\text{SYN NP[LID } \boxed{\langle 2 \rangle} \text{]} \\
\text{SEM} & \begin{bmatrix}
\text{INDEX} & x \\
\text{FRAMES} & \left\langle \ldots , \boxed{2}\begin{bmatrix} \textit{i-beans[secret]-fr} \\ \text{THEME} & x \end{bmatrix} \right\rangle
\end{bmatrix}
\end{bmatrix}
$$

$$
\begin{bmatrix}
\text{FORM} & \langle \text{the} \rangle \\
\text{SYN} & \text{DET} \\
\text{SEM} & \ldots
\end{bmatrix}
\quad
\begin{bmatrix}
\text{FORM} & \langle \text{beans} \rangle \\
\text{SYN} & \textit{noun}\text{[LID } \boxed{\langle 2 \rangle} \text{]} \\
\text{SEM} & \begin{bmatrix} \text{INDEX} & x \\ \text{FRAMES} & \langle \boxed{2} \rangle \end{bmatrix}
\end{bmatrix}
$$
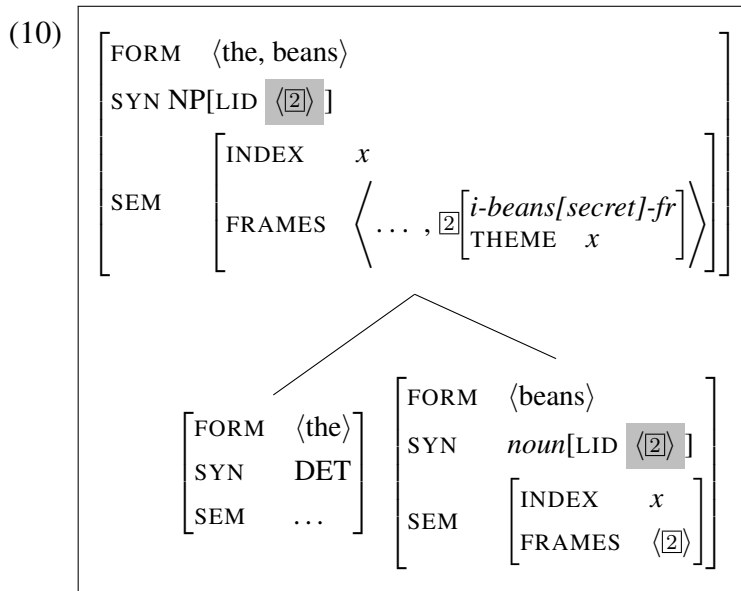
The basic mechanism for preventing idiom words from appearing in contexts where they are not lexically governed by the appropriate idiom predicator involves the feature VALENCE (VAL), Ordinary, non-idiom predicators are lexically specified as requiring all members of their VAL list to be nonidiomatic. By contrast, an idiomatic predicator like idiomatic *spill* requires a direct object that is idiomatic, i.e. one whose LID value (and hence that of its lexical head) contains a particular *i-frame*, as in (11b):[10]

---

[10]We follow the abbreviation conventions of Copestake et al. 2005 in (i) for the corresponding frame descriptions shown in (ii) (Here and throughout, UNDGR abbreviates UNDERGOER). Notice that we use s variables for quantifying over situations, which function as Davidsonian event arguments and omit label arguments except where necessary for label identification:

(11) a.
$$
\begin{bmatrix}
\textit{trans-v-lxm} \\[4pt]
\text{FORM} \quad \langle \text{spill} \rangle \\[4pt]
\text{SYN} \quad
\begin{bmatrix}
\text{VAL} \quad \left\langle \text{NP}_x[\text{LID } \langle \textit{c-frame} \rangle] \, , \, \text{NP}_y[\text{LID } \langle \textit{c-frame} \rangle] \right\rangle \\[4pt]
\text{LID} \quad \left\langle \boxed{1}\textit{spill(s,x,y)} \right\rangle
\end{bmatrix} \\[4pt]
\text{SEM} \quad
\begin{bmatrix}
\text{INDEX} \quad s \\
\text{FRAMES} \quad \langle \boxed{1} \rangle
\end{bmatrix}
\end{bmatrix}
$$

b.
$$
\begin{bmatrix}
\textit{trans-v-lxm} \\[4pt]
\text{FORM} \quad \langle \text{spill} \rangle \\[4pt]
\text{SYN} \quad
\begin{bmatrix}
\text{VAL} \quad \left\langle \text{NP}_x[\text{LID } \langle \textit{c-frame} \rangle] \, , \, \text{NP}_y[\text{LID } \langle \textit{i-beans[secret](y)} \rangle] \right\rangle \\[4pt]
\text{LID} \left\langle \boxed{1}\textit{spill[reveal](s,x,y)} \right\rangle
\end{bmatrix} \\[4pt]
\text{SEM} \quad
\begin{bmatrix}
\text{INDEX} \quad s \\
\text{FRAMES} \quad \langle \boxed{1} \rangle
\end{bmatrix}
\end{bmatrix}
$$

Because these predicators and their complements also pass their frames up to be part of the meaning of the phrases they license (in accordance with general principles of semantic composition), the idiomatic sense of the direct object *beans* will be part of the phrase's meaning only when the predicator *spill* has introduced the appropriate i-frame. A nonidiomatic object can only cooccur with predicators that fail to require an idiomatic object. Note that predicators like idiomatic *spill* are themselves identified via c-frames, which allows the phrases they project to occur in nonidiomatic contexts. The reader should bear in mind throughout that the basic mechanism allowing idiom words to occur just when governed by the appropriate lexical governor is that non-idiom predicators are lexically specified

---

(i)   *l: beans(x)*   *l:spill(s,x,y)*   *l:i-beans[secret](x)*   *l:spill[reveal](s,x,y)*

(ii)
$$
\begin{bmatrix}
\textit{beans-fr} \\
\text{LABEL} \quad l \\
\text{THEME} \quad x
\end{bmatrix}
\quad
\begin{bmatrix}
\textit{spill-fr} \\
\text{LABEL} \quad l \\
\text{SIT} \quad s \\
\text{ACTOR} \quad x \\
\text{UNDGR} \quad y
\end{bmatrix}
\quad
\begin{bmatrix}
\textit{i-beans[secret]-fr} \\
\text{LABEL} \quad l \\
\text{THEME} \quad x
\end{bmatrix}
\quad
\begin{bmatrix}
\textit{spill[reveal]-fr} \\
\text{LABEL} \quad l \\
\text{SIT} \quad s \\
\text{ACTOR} \quad x \\
\text{UNDGR} \quad y
\end{bmatrix}
$$

Representations of feature structures are enclosed in boxes; representations of descriptions of feature structures (including constructions) are not enclosed in boxes.

to accept only arguments whose LID value is canonical, as in (11a), which will always lead to a nonidiomatic VP like (12):

(12)

$$
\begin{bmatrix}
\text{FORM } \langle \text{spilled, the, beans} \rangle \\[4pt]
\text{SYN} \quad \text{VP} \begin{bmatrix} \text{CAT } [\text{LID } \langle \boxed{1} \rangle] \\ \text{VAL } \langle \boxed{5} \text{NP}_x \rangle \end{bmatrix} \\[10pt]
\text{SEM} \begin{bmatrix} \text{INDEX} \quad s \\ \text{FRAMES} \left\langle \begin{array}{l} \boxed{1} l_4 \text{:} spill(s,x,y), \ exist(s,l_1,l_2), \ l_1 \text{:} past(s), \\ l_2 \text{:} the(y,l_3,l_4), \ \boxed{2} l_3 \text{:} beans(y) \end{array} \right\rangle \end{bmatrix}
\end{bmatrix}
$$

$$
\begin{bmatrix}
\text{FORM} \langle \text{spilled} \rangle \\[4pt]
\text{SYN V} \begin{bmatrix} \text{CAT } [\text{LID } \langle \boxed{1} \rangle] \\ \text{VAL } \langle \boxed{5}, \boxed{6} \rangle \end{bmatrix} \\[10pt]
\text{SEM} \begin{bmatrix} \text{INDEX} \quad s \\ \text{FRAMES} \left\langle \begin{array}{l} \boxed{1}, \ exist(s,l_1,l_2), \\ l_1 \text{:} past(s) \end{array} \right\rangle \end{bmatrix}
\end{bmatrix}
\qquad
\boxed{6}
\begin{bmatrix}
\text{FORM} \langle \text{the, beans} \rangle \\[4pt]
\text{SYN NP} [\text{LID } \langle \boxed{2} \rangle] \\[4pt]
\text{SEM} \begin{bmatrix} \text{INDEX} \quad y \\ \text{FRAMES } \langle l_2 \text{:} the(y,l_3,l_4), \ \boxed{2} \rangle \end{bmatrix}
\end{bmatrix}
$$

And the idiomatic (11b) will project structures like (13), where the semantics includes the direct object's *i-frame*:

(13)

$$\begin{bmatrix} \text{FORM} & \langle \text{spilled, the, beans} \rangle \\ \text{SYN} & \text{VP} \begin{bmatrix} \text{CAT} & [\text{LID} \ \langle \boxed{1} \rangle] \\ \text{VAL} & \langle \boxed{5}\text{NP}_x \rangle \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & s \\ \text{FRAMES} & \left\langle \begin{array}{l} \boxed{1}l_4\text{:}spill[reveal](s,x,y), \ exist(s,l_1,l_2), \ l_1\text{:}past(s), \\ l_2\text{:}the(y,l_3,l_4), \ \boxed{2}l_3\text{:}i\text{-}beans[secret](y) \end{array} \right\rangle \end{bmatrix} \end{bmatrix}$$

$$\begin{bmatrix} \text{FORM} \langle \text{spilled} \rangle \\ \text{SYN} \ \text{V} \begin{bmatrix} \text{CAT} & [\text{LID} \ \langle \boxed{1} \rangle] \\ \text{VAL} & \langle \boxed{5}, \ \boxed{6} \rangle \end{bmatrix} \\ \text{SEM} \begin{bmatrix} \text{INDEX} & s \\ \text{FRAMES} & \left\langle \begin{array}{l} \boxed{1}, \ exist(s,l_1,l_2), \\ l_1\text{:}past(s) \end{array} \right\rangle \end{bmatrix} \end{bmatrix}$$
$$\boxed{6} \begin{bmatrix} \text{FORM} \langle \text{the, beans} \rangle \\ \text{SYN} \ \text{NP}[\text{LID} \ \langle \boxed{2} \rangle] \\ \text{SEM} \begin{bmatrix} \text{INDEX} & y \\ \text{FRAMES} & \langle l_2\text{:}the(y,l_3,l_4), \ \boxed{2} \rangle \end{bmatrix} \end{bmatrix}$$

Our analysis thus treats *Kim spilled the beans* as ambiguous. (11a) will give rise to an S whose FRAMES list is as shown in (14a) and (11b) will likewise project an S whose FRAMES list is (14b):

(14) a. $l_0\text{:}\exists(s,l_1,l_2), \ l_1\text{:}past(s), \ l_2\text{:}the(y,l_3,l_4), \ l_3\text{:}beans(y), \ l_4\text{:}spill(s,Kim,y)$

 b. $l_0\text{:}\exists(s,l_1,l_2), \ l_1\text{:}past(s), \ l_2\text{:}the(y,l_3,l_4), \ l_3\text{:}i\text{-}beans[secret](y),$
     $l_4\text{:}spill[reveal](s,Kim,y)$

The ls here are just labels (see Copestake et al. 2005) that can be eliminated by replacing an argument l by the formula labelled by l, as shown in (15):

(15) a. (14a) = $\exists(s,past(s), \ the(y, beans(y), spill(s,Kim,y)))$

 b. (14b) = $\exists(s,past(s), \ the(y, i\text{-}beans[secret](y), spill[reveal](s,Kim,y)))$

(15b) is the idiomatic reading, which can be read as 'there is some past situation in which there is a contextually provided secret that Kim reveals'.

3 The Analysis of Semi-Fixed Expressions

We return now to the question of how an idiom like *kick the bucket* should be analyzed. The first question that arises is whether to attribute the meaning 'die' to *kick* alone or to the full VP *kick(s/ed/ing) the bucket*.[11]

One might propose to analyze *kick the bucket* by positing two lexemes, an idiom-predicator verb *kick* meaning 'die' (punctually) which requires as object a meaningless noun *bucket* marked by *the*.[12] On this account, the absence of Passive stems from a restriction on the VERBFORM (VF) value of *kick*. Such an analysis, apart from being stipulative, does not yet deal with the entire problem raised here, since Passive is not the only construction where this idiom is blocked. No idiomatic readings are available for any of the examples in (16), for example:

(16) a. Which (gunman's) bucket did he kick?

    b. What he kicked was the bucket.

    c. It was the bucket that he kicked.

    d. In Kim's case, the bucket kicked easily.

This objection might be countered by observing that Questions, *WH*-clefts, *It*-clefts and Middle all impose semantic or pragmatic constraints on the (displaced) direct object of the verb, which would make these constructions incompatible with any analysis of idiomatic *kicked*, that required a semantically vacuous direct object. It seems unlikely that the failure of Passive can be explained in these terms,

---

[11]Presumably the absence of a progressive usage, as in (i) (cf. (ii) and (iii)) is explained by punctuality of meaning.

  (i) *He was kicking the bucket when she walked in.

 (ii) Erasmus seized him by the throat, threw him down, and giving him a violent kick in the neck, a rattle in the throat announced that he was giving up the ghost.

(iii) Lowe and Cushing sank bleeding to the basket floor as the balloon was giving up the ghost, its descent accelerating as the weight of its fabric overpowered the …

Example (iv) shows that the gerund form is possible for *kick the bucket*, so if gerund and present participle are taken to employ the same morphological object, the progressive cannot be blocked morphologically.

 (iv) I think the correlation between her kicking the bucket and the surgery is a daft point.

[12]How this mark results in the NP headed by *bucket* being determined by *the* is explained below.

however, as there are meaningless elements that can appear as Passive subjects, making it unclear what the semantic conflict between idiomatic *kick the bucket* and Passive might be:

(17) a. There was believed to be another worker at the site besides the neighbors who witnessed the incident.

   b. It was rumored that Great Britain, in apparent violation of the terms of the Clayton-Bulwer treaty, had taken possession of certain islands in the Bay of Honduras.

The absence of Passive realizations of *kick the bucket* and the like must be understood in terms of the larger problem of predicting when passivization is possible. There are several classes of verbs that either inherently or in specific uses cooccur with an immediately following NP complement which cannot be realized as passive subjects. We will call these *pseudotransitive* verbs and posit a type *pseudotransitive-verb-lexeme* (*pseudo-tv-lxm*). Examples include the class of verbs that take measure objects such as *cost, weigh,* and *measure*. The listemes for these lexemes will specify membership in the type *pseudo-tv-lxm*, specifically in the subtype *measure-verb-lexeme* (*meas-v-lxm*). There are also verbs that are not inherently of this type, but which fail to passivize just when they take measure NP complements; example are *reach*, *attain*, *stretch*, and so on:

(18) a. As is typical with hand dug sites, a depth of aproximately 9 feet was reached by those that dug.

   b.*As is typical with hand dug sites, aproximately 9 feet was reached by those that dug.

   c. The Pocomoke River stretches approximately 66 miles (106 km) from southern Delaware through southeastern Maryland in the United States.

   d.*Approximately 66 miles (106 km) is stretched by the Pocomoke River...

We thus posit a lexical construction (analogous to a lexical rule in many approaches) that converts transitive verb lexemes to intransitive verbs requiring a measure phrase instead of a direct object NP. Since the Passive Construction builds a past participle from a transitive verb lexeme, this analysis successfully predicts the contrasts shown in (18).

Another class of verbs subject to an analysis of this kind is illustrated in (19):

(19) a. Luxury consumers shop Neiman Marcus Christmas Book online.

    b.*Neiman Marcus Christmas Book is shopped by luxury Consumers online.

    c. …but the next day she flew United to France and said the manicotti on the airplane was better…

    d.*but the next day United was flown (by her) to France and she said the manicotti on the airplane was better…[13]

    e. Ski Vermont!

    f.*Vermont is skied a lot.

These uses have been analyzed within the Relational/Arc-Pair Grammar tradition as 'advancements', e.g. as 'locative-to-2' or 'instrumental-to-2' advancements. Our analysis differs in denying direct object ('2') status to the postverbal NPs in (19). We again assume an analysis in terms of a lexical construction, one whose 'output' (MOTHER) is a verb lexeme, but not a transitive verb lexeme. This will guarantee that a verb like those in (19) has no corresponding passive participle.

The *kick* of *kick the bucket* belongs to yet another class of pseudotransitives: those with pleonastic pseudo-objects:

(20) a.*The boys were afraid that when the principal found out his stack/top would be blown.

    b.*Sooner or later a cropper is come by every miscreant.

    c.*The dust was bitten by several members of the posse.

    d.*Uncle will never be said by Julian Assange.

We thus posit a type *pleonastic-pseudo-object-verb-lexeme* (*pleo-pso-v-lxm*) to which all the idiomatic verbs in (20) belong. As before, the lexical Passive Construction specifies that its daughter is of the type *trans-v-lxm* (and its mother of type *passive-participle-word* (*pass-part-word*)). Hence none of the pseudo-transitive verbs we have discussed, including those classified as *pleo-pso-v-lxm*) will passivize. The non-passivizability of idiomatic *kick* and lexemes like it is thus accounted for by the semantically motivated partition of the lexicon sketched in (21) and its interaction with general properties of the Passive construction.

---

[13]Charles Fillmore, personal communication.

(21)

$$\text{pseudo-tv-lxm}$$

$$\textit{meas-v-lxm} \quad \textit{loc-pso-v-lxm} \quad \textit{pleo-pso-v-lxm} \quad \ldots$$
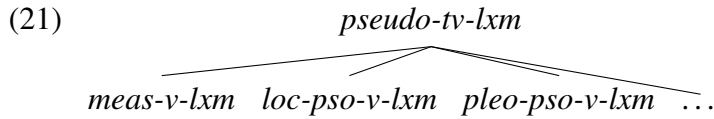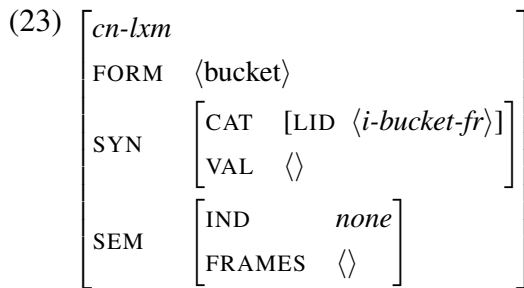
This strictly lexical analysis of *kick the bucket* accounts in addition for both the inflectional freedom of the head *kick* and the fact that the *bucket* NP can be interrupted by metalinguistic operators like *proverbial*, *metaphorical*, etc., as in (22):

(22) a. I feel like I've seen so much of the world already, but there's still a lot I want to do before I kick the proverbial bucket. . .

    b. My Kindle kicked the metaphorical bucket on the 25th.

Since idiom *kick* specifies about its NP complement only that it is an NP headed by the idiom word *bucket*, the insertion of epithets or semantically external modifiers in examples like those in (22) causes no problem, while semantic modification of the idiomatic word *bucket* is precluded by that noun's lexically specified meaninglessness.

We now turn to our analysis of the idiomatic common noun *bucket* in SBCG terms. This involves a listeme (lexical entry) that is like the one for nonidiomatic *bucket*, except that (1) the FRAMES list is empty and (2) the frame contained in the LID value is of type *i-bucket-fr*:

(23)
$$
\begin{bmatrix}
\textit{cn-lxm} \\
\text{FORM} \quad \langle \text{bucket} \rangle \\
\text{SYN} \quad
\begin{bmatrix}
\text{CAT} \quad [\text{LID} \quad \langle \textit{i-bucket-fr} \rangle] \\
\text{VAL} \quad \langle \rangle
\end{bmatrix} \\
\text{SEM} \quad
\begin{bmatrix}
\text{IND} \quad \textit{none} \\
\text{FRAMES} \quad \langle \rangle
\end{bmatrix}
\end{bmatrix}
$$

The label *cn-lxm* indicates that idiomatic *bucket* is a common noun lexeme, a subtype of *lexeme*. The first syntactic specification requires that the LID feature have the value $\langle$*i-bucket-fr*$\rangle$, the list containing just a feature structure of type *i-bucket-fr*. As indicated above, the type hierarchy recognizes two kinds of frames: canonical frames, subtypes of *c-frame* and idiomatic frames, subtypes of *i-frame*. In the case of meaningless idiomatic lexemes, those with an empty FRAMES list and *none* as their INDEX value, like idiomatic *bucket*, the LID value is a semantically

14

inert *i-frame* that simply identifies the lexeme for purposes of selection by the relevant predicator. As before, since CAT features percolate in headed constructs from the head lexeme to higher projections, the identity of idiomatic *bucket* as the head of a noun phrase like *the bucket* or *the proverbial bucket* will be visible to idiomatic *kick*.

Before turning our attention to idiomatic *kick*, let us consider the construction that licenses the NP headed by idiomatic *bucket*. The Head-Functor Construction, formulated in (24), licenses constructs of type *hd-func-cxt*, a subtype of *headed-construct*; a *headed-construct* specifies one of its daughters as the head daughter and further specifies that the head daughter's syntactic category (SYN|CAT) value is shared by the mother, encoding the principle of head feature percolation.[14]

(24) **Head-Functor Construction: (↑*headed-cxt*):**

$$
\textit{hd-func-cxt} \Rightarrow
\begin{bmatrix}
\text{MTR} & [\text{SYN } X![\text{MRKG } M\,]] \\[4pt]
\text{DTRS} & \left\langle
\begin{bmatrix}
\text{SYN} &
\begin{bmatrix}
\text{CAT } [\text{SELECT } Y\,] \\
\text{MRKG } M
\end{bmatrix}
\end{bmatrix}, \; Y{:}[\text{SYN } X]
\right\rangle \\[4pt]
\text{HDDTR} & Y
\end{bmatrix}
$$

In our discussion of *spill the beans*, we already sketched a construct licensed by this construction; see (10) above.

We notice in the title line of (24) the notation ↑*headed-cxt*, which is to be read as '*headed-construct* is an immediate supertype of *head-functor-construct*'. Constructs play an important role in SBCG: they are like local trees licensed in a Context-Free Grammar; similarly, the combinatoric constructions that license these constructs are analogous to Context-Free rewrite rules. Since a construct is modeled as a feature structure, it too is specified in AVM notation: its MTR (MOTHER) feature's value is a sign and its DTRS (DAUGHTERS) feature's value is a non-empty list of signs.

Just as 'VP → VP ADV' places constraints on a type of head-modifier structures in a CFG, the Head-Functor Construction constrains a class of headed structures in a SBCG. Note that the constraints in both cases are local, making reference only to two levels of structure. The notation for SBCG constructions presents

---

[14]Based on the analyses of Allegranza (2007), Van Eynde (2006, 2007), Sag (2012), and Kay and Sag (2012), we posit a SELECT feature, according to which specifiers, markers, and modifiers select their heads, in effect collapsing the categories of specifier, modifier, and marker.

a double arrow with the type name of a construct on its left, here *head-functor-construct* (*hd-func-cxt*) and on its right an AVM expressing the defining properties of the class of constructs named on the left. Signs are thus either licensed by lexical entries, *listemes*, or else they are 'constructionally licensed' as the mother of well-formed constructs.[15]

The defining constraint on the class of *hd-func-cxt*s says something about the mother, the daughters and the head-daughter. The notation in the MTR's value says that the SYN value is the same as that associated elsewhere with the variable X, except that the local value has the marking value indicated by the variable M. (The latter, 'except' clause is the interpretation of the '!' following the X.) The DTRS value specifies a list of two daughters. Taking the second daughter first, its SYN value is tagged X, which, when taken together with the occurrence of 'X !' in the MTR's SYN value, says that the MTR's syntax value is that of its second daughter except for the specification [MRKG M], indicated by '!'. The second daughter is tagged Y, which identifies it as the head daughter (i.e. the value of the HDDTR feature just below). Turning now to the first daughter (the functor), we note that its MRKG value is M, and its SYN|CAT|SELECT value is Y, the variable also assigned to the head daughter. The former specifies that in constructs of type *hd-func-cxt* the mother inherits its MARKING value from the non-head daughter (while it inherits the rest of its syntax from the head daughter, as we have seen). The latter indicates that the 'functor' daughter selects the head daughter. In the case of idiomatic *kick the bucket*, this specification will allow *the* to select a nominal phrase headed by idiomatic *bucket*.[16] The Head-Functor Construction licenses the combination of non-quantifier *the* and idiomatic *bucket* (or a *bucket*-headed phrase with appropriate modification) into an NP obeying the constraints that idiomatic *kick* will require of its pseudo-object.

Before considering the special lexical entry for idiomatic *kick* we consider the construction that licenses phrases consisting of predicators with their non-subject complements, i.e. the construction that will enable *kick* to combine with *the. . . bucket*. The Predicational Head-Complement Construction licenses head-

---

[15]There are also unilevel constructions in the lexicon, called *lexical class constructions*. We are not concerned with them in this paper. See Sag 2012.

[16]We assume that English has a meaningless *the*, which occurs in this and several other idioms, such as *trip the light fantastic*, *drop the ball*, *bite the dust*, *fly the coop*, and *buy the farm*. Throughout our discussion of *kick the bucket* the *the* we have in mind is this meaningless *the*. It has the same syntactic properties as quantifier *the*; these include selecting a common noun phrase as sister and carrying the marking value *the*, which in the NP built by the Head Functor Construction will be percolated to the mother.

complement phrases—VPs, PPs, APs, and Common Noun Phrases (CNPs)—in which all non-subject complement requirements (non-subject *valents*) of a predicator are realized as sisters to the head. As in the Head-Functor-Construction, we note in the title line of (25) that this construction also licenses a subclass of headed constructs:

(25) **Predicational Head-Complement Construction (↑*headed-cxt*):**

$$
\textit{pred-hd-comp-cxt} \;\Rightarrow\; \begin{bmatrix} \text{MTR} & [\text{SYN } X \:!\: [\text{VAL } \langle Y \rangle]] \\ \text{DTRS} & \langle\, Z \,\rangle \;\oplus\; \textit{L:nelist} \\ \text{HDDTR} & Z\text{:} \begin{bmatrix} \textit{word} \\ \text{SYN } X : \begin{bmatrix} \text{CAT} & [\text{XARG } Y] \\ \text{VAL} & \langle Y \rangle \oplus L \end{bmatrix} \end{bmatrix} \end{bmatrix}
$$

Our earlier discussion exhibited two constructs, (12) and (13), that are licensed by (25).

As in the Head-Functor Construction (24), MTR, DTRS, and HDDTR features are all constrained by (25). Looking first at the DTRS value, we note that it is a composite list, composed of a singleton list containing a sign tagged Z, which we see below is identified with the head daughter, followed by a non-empty list, tagged L. Viewed as a local tree this class of constructs is characterized by a list of one or more daughters of which the head daughter is the first. Looking down now at the head daughter, tagged Z, we see that it is of type *word* and that its SYN value, tagged X, has specifications for CAT and VAL features. The CAT feature specifies that its external argument (XARG) is tagged Y—which we see again in the MTR's SYN|VAL list. In general, the XARG feature, as a CAT feature, is projected upwards from the lexical head in all headed constructs, making visible at all 'bar' levels the subject requirement of a lexical predicator. Normally, the XARG value is identified with the first element of the VAL list, which lists the unsaturated argument requirements of a predicator.[17] The VAL list of the head daughter consists of a singleton list containing the XARG (subject-to-be), tagged Y, which is identified with the unique member of the VAL list of the mother, followed by a non-empty list tagged L, whose members, we see in the DTRS value, are identified with the non-head daughters. The Predicational Head-Complement

---

[17]The VAL list corresponds fairly closely with the SUBCAT list of Pollard and Sag (1994). For further discussion of the XARG feature, especially regarding its relation to issues of locality and long distance dependency, see Sag 2010b.

Construction will license a VP headed by idiomatic *kick* followed by an NP headed by idiomatic *bucket*.

We have seen how the Head-Functor Construction licenses an NP headed by idiomatic *bucket*, selected by *the*, and possibly containing semantically external modification: e.g. *proverbial* or *metaphorical*. We have also discussed the construction that assembles verb phrases, among others, and their non-subject complements. It remains to complete the picture by considering the listeme for idiomatic *kick*, which is given in (26):

(26)
$$
\begin{bmatrix}
\textit{pleo-pso-v-lxm} \\
\text{FORM} \quad \langle \text{kick} \rangle \\
\text{SYN} \quad \begin{bmatrix} \text{VAL} \quad \left\langle \text{NP}_i \,, \text{NP}\begin{bmatrix} \text{SYN} \begin{bmatrix} \text{CAT}|\text{LID} \ \langle \textit{i-bucket-fr} \rangle \\ \text{MKG} \quad \textit{the} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix} \\
\text{SEM} \quad \begin{bmatrix} \text{FRAMES} \quad \left\langle \begin{bmatrix} \textit{die-punctually-fr} \\ \text{DECEDENT} \ i \end{bmatrix} \right\rangle \end{bmatrix}
\end{bmatrix}
$$

As shown in (26), idiomatic *kick* is a pseudo-transitive lexeme whose pseudo-object is semantically empty and whose base morphological form is *kick*. This lexeme has two valence requirements. The first is for a potential subject NP with index i. The second is for an NP identified in terms of *i-bucket-fr*, an idiomatic frame that never contributes to the sentence's meaning. The semantics of the lexeme *kick* consists of the predicate 'die punctually' whose argument is identified with the index of its potential subject. The second valent requirement will ensure that this verb projects a VP containing a single NP. whose head lexeme is idiomatic *bucket* and whose functor ('specifier') is meaningless *the*. The resulting VP will express a predicate 'die punctually' that is 'looking for' a subject NP to provide the DECEDENT argument to complete the predication.

This concludes our analysis of *kick the bucket*. Notice that the present approach eliminates the need for special devices for the analysis of semi-fixed expressions. Instead, they are analyzed in the same way as syntactically flexible expressions, leaving their restricted properties to be explained in terms of the idiosyncrasies of their listemes in interaction with ordinary phrasal constructions.

4 Super-Flexible Idioms

The idioms we have examined so far exhibit varying degrees of flexibility. In this section we explore the analysis of idioms whose variability includes more than multiple morphological realizations or the positional variation of a single idiom word. We first examine *side the bread is buttered on* and *What's X doing Y?*, which interact with *wh*-constructions in interesting ways, yet exhibit certain differences which, we claim, can be simply treated in terms of diverse lexical constraints and their interaction with the constraints imposed by English combinatorial constructions and other listemes.

Consider first the sentences in (27):

(27) a. Yes, a little line drawn in the sand to let Bibi know which side his bread is buttered on.

b. I know which side my bread is buttered on.

c. Google should know which side its bread is buttered on.

d. And don't think for a minute that the "scientific community", especially the Al Gore global warming community, doesn't know what side their bread is buttered on.

We have not found any linguistic analyses of this idiom. Popular accounts usually say—or more often imply—that it has the following properties:

A. The matrix verb is *know*, possibly negated.

B. The specifier of *bread* is a genitive pronoun.

C. The pronominal specifier of *bread* is bound to the subject of *know*.

D. The NP headed by *side* must cooccur with an interrogative *wh*-word determiner.

E. The *bread-is-buttered-on* phrase is an embedded interrogative clause.

F. The immediate governor of the phrase headed by *buttered* is BE, presumably the passive auxiliary.

G. The preposition *on* is stranded.

19

The next section may be viewed as a methodological exercise. The idiom is more abstract than suggested by the above picture. In fact none of the properties (A–G) is a necessary one. Often, when one notices a recurring grammatical pattern and begins to investigate its constructional status, one finds that the truly idiomatic part—the irreducible stuff that is unpredictable from anything else in the grammar—is more abstract than the pattern one originally recognized. Following this reduction of the scope of the syntactic investigation, we will look at the semantics of the idiom. Then a somewhat more formal analysis of the construction will be presented, in an effort to persuade the reader of the analytic versatility of Sign-Based Construction Grammar.

4.1 Limiting the Scope of the Bread-Buttering Idiomaticity

A. It is immediately apparent that many verbs other than *know* can serve as governors in this pattern.

(28) a. Come on, this guy figured out which side his bread is buttered on …

b. Looking at the headlines, I notice that US Senators from both parties are paying careful attention to the side their bread's buttered on.[18]

It appears that (at least) any verb that can take an embedded interrogative complement can appear as the governing verb in this pattern.

---

[18]Some readers may not consider examples like this one, in which the specifier of *side* is not a *WH*-phrase, legitimate examples of ordinary English, but the web contains too many examples of this type to consider them all word play:

(i) He's a so called republican because that's the side his bread's always been buttered on and the Dems probably nauseated him.

(ii) The Side on Which My Bread is Buttered [Blog title]

(iii) In this primer of power, we have all the crudities of the youngster who has as yet inadequate conceptions of the side on which his bread is buttered. …

(iv) The side the bread is buttered on [headline]

(vi) Mar 3, 2008. if these guys step out of line I'm sure the side their bread is buttered on will become apparent. In the American retail world, …

(vii) The side one's bread is buttered on tends to affect one's credibility.

B. The specifier of *bread* need not be a pronoun. It may be a genitive noun phrase, either proper or common, or it may be the definite article *the*. The constraint seems to be only that it must be definite:

(29) a. In his pursuit to find out which side Benji's bread is buttered on, Andy J,...

    b. Why is this drivel on Massey's site? Makes one wonder who's side WSAZ's bread is buttered on!

    c. I chose the super closeup to convey the fact that we don't know which side the guy's bread is buttered on. Eg. Goodie or baddie.

    d. pictures of Arafat as if he was a hero when even Arafat knew which side the bread was buttered on.[19]

C. When the specifier of *bread* is a pronoun it is not necessarily bound to a higher subject.

(30) a. I guess we know which side your bread is buttered on.

    b. I know what side your bread is buttered on, I'm in, I'm ONE OF YOU.

    c. I know which side Haseo's bread is buttered on.

    d. He had the manners and listening attitude down pat—but you know which side his bread is buttered on.

    e. I'm pretty sure that we know which side Michelle's bread is buttered on.

---

[19]More examples of this type for those inclined to reject these, again selected from many others attested on the web:

  (i) If this is not proof as to which side the MSM'S bread is buttered, then I don't know what is!

  (ii) It is now clear, though, on which side the Global Times's bread is buttered.

  (iii) LOL I think we know which side the umpire's bread was buttered!

  (iv) a little checking showed which side the military's bread was buttered on and the side they'd ...

  (v) Maria knows on which side the family's bread is buttered, and realizes that without the income generated by Vermeer's oils, the household would be thrown into the street.

  (vi) Cameron knows which side the UK's bread is buttered on.

D. The noun phrase headed by *side* need not contain a *WH*-word:[20]

(31) a. Gmail allows you to use an email client, but they restrict some of the goodies to web access, because that is the side their bread is buttered on.

    b. but if these guys step out of line I'm sure the side their bread is buttered on will become apparent.

E. The pattern may occur without all or any part of it constituting an embedded question—(31a) and (32), for example:

(32) a. That's the side Buckley's bread is buttered on.

    b. Nvidia know that this is the side their bread is buttered on …

    c. The sage whom thus we celebrate Was bound to pray for Church and State; And piously his prayers he mutter'd, For on that side his bread was butter'd.

    d. Married into the nation's most prominent Democrat family, California Gov. Arnold Schwarzenegger flopped to the GOP side that his bread is buttered on.

F. A verbal governor of the past participle *buttered* need not be the passive auxiliary *be*. The *get*-passive is also possible, although admittedly rare:

(33) a. Rep. Bachmann said. "They are incredulous about the possibility of losing their majority and they know which side their bread gets buttered on and ACORN is their friend."

    b. Beck knows which side his bread gets buttered on.

    c. This story still uses liberal connotations and loaded terms, though, so you can tell which side their bread gets buttered on. (punctuation added)

    d. TRUE - Stafford will be back at QB and Detroit has made some good moves this off season. They know what side their bread gets buttered on.

---

[20]See also footnote 18.

e. It's no surprise then that they would sugar coat or down play anything which makes their benefactors look bad, they know what side their bread gets buttered on.[21]

G. Perhaps surprisingly, Pied Piping tokens of this idiom are about as frequent on the web as prepositional stranding ones. In actual Google hits *on ∗ side ∗ bread ∗ buttered* scores 480 versus 490 hits for *side ∗ bread ∗ buttered on*. A few examples of Pied Piping:

(34) a. They know on which side their bread is buttered and will adapt quite quickly to what they see as the potential winning side.

b. Wendy knows on which side her bread is buttered.

c. Still, few Costa Ricans have anything bad to say about their country's popularity as a destination—perhaps simply because they know on which side their bread's buttered.

d. Salespeople know on which side their bread's buttered.

The contraction of *bread's* in the last two examples suggests that the Pied Piping version of the idiom is not restricted to formal contexts.

4.2 The Analysis of *the side the bread is buttered on*

As noted, our analysis of this idiom consists entirely of lexical entries ('listemes' in SBCG parlance). The properties of these listemes interact with the general phrasal constructions of the grammar, such as the Predicative Head-Complement Construction, which licenses VPs and non-case-marking PPs, and the Head-Functor Construction, which licenses binary branching NPs (*inter alia*), to determine the syntactic privileges of occurrence of the idiom and to compose the semantics of each token.

---

[21]The use of *get*, rather than *be*, in this idiom is relatively rare. Google yields only 50 true hits for *their bread gets buttered on* versus 498 for *their bread is buttered on*. We have not attempted to compare this ratio of roughly 1/10 to the relative frequencies of the *get-* and *be*-passives. A case could be made to count examples like those in (33) as word play or other forms of nonce extension of the grammar, thus excluding them from the data on which analysis of the construction is based. Judgment calls of this kind are seemingly unavoidable in grammatical research, even when based on corpus data.

We gloss the meaning of the *side someone's bread is buttered on* idiom as something like the '**location** (*side*) at which [someone's] **needs or desires** (*bread*) are **satisfied** (*buttered*)', where the key frame (relation, predicate) is that of satisfaction or gratification and its arguments are a THEME, the 'needs', and a LOCATION. Someone who knows 'the *side* their *bread* is *buttered* on', knows the **place** where their **needs** are **satisfied**.[22] The FRAMES list of the semantically key *buttered* listeme is given in (35):

(35)

$$
\left[
\text{FRAMES} \quad \left\langle
\begin{bmatrix}
\textit{buttered[satisfy]-fr} \\
\text{LOCATION} \quad y \\
\text{THEME} \quad x
\end{bmatrix}
\right\rangle
\right]
$$

The *buttered[satisfy]-fr* is a canonical frame (*c-frame*); since *buttered* is the governing lexeme of the idiom, the phrase it projects (and to which it contributes its LID value) occurs in canonical contexts. The idiom says something about where someone's needs are satisfied, but nothing explicitly about who or what they are satisfied by. When we assert—or especially question or deny—someone's knowing the side their bread is buttered on, we may or may not have a particular benefactor in mind. We will see below that the value of the THEME feature of the *buttered[satisfy]-fr* is identified with the index of the *bread* constituent and the value of the LOCATION feature is identified with the index of the NP headed by *side*, which is also the index of the PP headed by *on*.

The syntactic analysis of the idiom consists entirely of lexical entries (listemes), exploiting the original NSW proposal that semantic and syntactic composition of idioms go hand in hand. The syntactic possibilities and constraints follow from the properties of the listemes for idiomatic *buttered, bread, on*, and *side* in interaction with various phrasal constructions of the grammar. We posit an idiomatic passive participle *buttered* 'satisfied' with two arguments: an NP headed by the idiomatic noun *bread* 'needs/desires' and a prepositional constituent that resolves either to a PP headed by an idiom preposition *on* or to the preposition *on* itself, depending on whether a given token employs the usual mechanisms of Pied Piping or Stranding (See Sag 2010a, 2012). A feature structure modeling the idiomatic participle *buttered* is sketched in Figure 1.

The idiomatic preposition *on* has an object argument headed by idiomatic *side* (See (36)). As noted above, the *side* argument denotes the location where needs

---

[22]We can imagine alternative semantic parses of the idiom and expect inter-speaker variation in this and similar cases.

$$
\begin{bmatrix}
\textit{word} \\
\text{FORM} \quad \langle\,\text{buttered}\,\rangle \\
\text{ARG-ST} \quad \left\langle\, X{:}NP_y\begin{bmatrix}\text{LID} & \langle\textit{i-bread[needs]-fr}\rangle \\ \text{MRKG} & \textit{def}\end{bmatrix},\ PP_x[\text{LID}\ \langle\textit{i-side[loc]-fr}\rangle]\,\right\rangle \\
\text{SYN} \quad \begin{bmatrix}
\text{CAT} & \begin{bmatrix}\textit{verb} \\ \text{LID} & \langle\boxed{1}\rangle \\ \text{VFORM} & \textit{pas}\end{bmatrix} \\
\text{VAL} & \langle X,\ldots\rangle \\
\text{GAP} & \langle NomP_x\rangle
\end{bmatrix} \\
\text{SEM} \quad \begin{bmatrix}\text{IND} & s \\ \text{FRAMES} & \langle\,\boxed{1}\textit{buttered[satisfy](s,x,y)}\,\rangle\end{bmatrix}
\end{bmatrix}
$$

Figure 1: The Idiomatic Participle *buttered*

are satisfied (See (38)). The *on* preposition identifies its index with that of its object, and hence refers to the 'location' at which 'needs' are 'satisfied'. The NP headed by *bread* denotes the 'needs' (see (37)). The first argument of *buttered*, is an NP lexically headed by *bread* and, as demonstrated in the data section, must be definitely determined, by either a genitive NP or *the*. The *buttered* listeme also has a non-empty GAP list containing a nominal phrase (NP or PP) coindexed with the *on* constituent (and hence with the *side* constituent). Thus either the *on* PP or its NP *side* object will normally appear in a filler position, licensed by some filler-head construction. The listemes for *buttered* and *on* are thus formulated with sufficient abstraction to allow the filler phrase to resolve to either the preposition-stranded or pied-piped form ([*Remember*] *which side the bread is buttered on.* vs. [*Remember*] *on which side the bread is buttered.*) Consequently, the expressed complement of *buttered* may be either the stranded preposition *on* or nothing (assuming a traceless theory of gaps like those in Sag 2010a or Chaves 2012).

The LID value of *buttered[satisfy]-fr* (a *c-frame*) identifies it as a canonical lexeme; the FORM value is $\langle$buttered$\rangle$. The ARG-ST specifies two arguments. The first argument, tagged X, is required to be an NP headed by idiomatic *bread*, whose index is y, and which is marked as definite. This argument is identified in the semantics with the THEME of the *buttered[satisfy]-fr*, i.e. with the needs to

be satisfied. Unless an argument (member of the ARG-ST list) of a predicate is extracted, it is identified with the corresponding element of the VAL list. Hence *buttered* is specified as [VAL $\langle X, \ldots \rangle$].[23]

The second member of the ARG-ST list is a prepositional phrase whose LID identifies it as being headed by idiomatic *on* and which is indexed x. The GAP list contains a nominal phrase (that is, either an NP or a PP) coindexed with the *on* PP of the ARG-ST list. If that NomP is resolved to the whole PP (i.e. if the PP is pied-piped), the VAL list will have no member reflecting the second member of ARG-ST list. If only the NP object of *on* appears on the GAP list, then the bare preposition *on* will be stranded and so appear on the VAL list.

The listeme for idiomatic *on* is given in (36). It requires an idiomatic *side-*headed NP object, has no semantics of its own, but identifies its index and its LID with that of its object:

(36)
$$\begin{bmatrix} \textit{trans-p-lxm} \\ \text{FORM} & \langle \text{on} \rangle \\ \text{SYN} & [\text{CAT } [\text{LID } \langle Z \rangle]] \\ \text{ARG-ST} & \langle \text{NP}_y[\text{LID } \langle Z\text{:}\textit{i-side[loc]-fr}\rangle] \rangle \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & y \\ \text{FRAMES} & \langle \rangle \end{bmatrix} \end{bmatrix}$$

For completeness, the listemes for idiomatic *bread* and *side*, which present no surprises, are given in (37) and (38):

(37)
$$\begin{bmatrix} \textit{cn-lxm} \\ \text{FORM} & \langle \text{bread} \rangle \\ \text{SYN} & [\text{CAT } [\text{LID } \langle Z \rangle]] \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & y \\ \text{FRAMES} & \langle Z\text{:}\textit{i-bread[needs](y)}\rangle \end{bmatrix} \end{bmatrix}$$

---

[23]This is accomplished not by redundant stipulation, but by general principles of argument realization, e.g. the Argument Realization Principle discussed in Sag et al. 2003, which also deals with "null instantiations" (Fillmore 1986) as nonvalent arguments.

(38)
$$
\begin{bmatrix}
\textit{cn-lxm} \\
\text{FORM} \quad \langle \text{side} \rangle \\
\text{SYN} \quad [\text{CAT } [\text{LID } \langle Z \rangle]] \\
\text{SEM} \quad
\begin{bmatrix}
\text{INDEX} \quad x \\
\text{FRAMES} \quad \langle \textit{Z:i-side[loc](x)} \rangle
\end{bmatrix}
\end{bmatrix}
$$

Let us illustrate a clause built by the interaction of the listemes constituting this idiom with the relevant phrase-building constructions of English:

(39) [He finally figured out] what side the bread is buttered on.
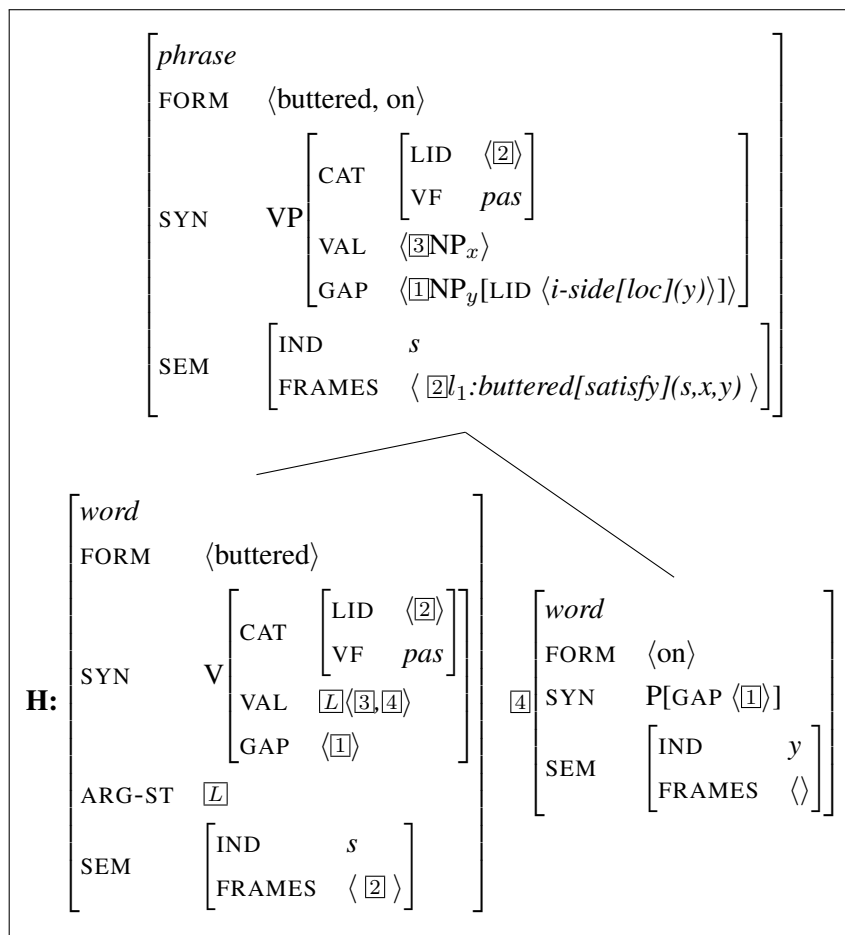
Figure 2: A Predicational VP Construct: *buttered on*

In Figure 2 we see that the LID of the passive participle *buttered* is identified with the content of the FRAMES value, via the tag ②.[24] And, since LID is a CAT feature, it reappears as the corresponding feature of the mother phrase *buttered on*. The VALENCE list of *buttered* is identified, via the tag Ⓛ, with its ARG-ST value, since there is no extraction or null argument instantiation for *buttered*. The VAL list contains the tags ③ and ④, the second of which tags the *on* constituent, which is the right sister to *buttered*. Since the first valent, ③ is not realized as a sister, it reappears in the VAL list of the mother, *buttered on* (uncanceled valents are passed up to an expression's mother, as in Categorial Grammar). Although there is no extraction or null instantiation of arguments of *buttered*, there is nonetheless a non-empty GAP list, since there is extraction of the object of the complement sister *on*.[25] Since a mother phrase inherits the gaps of its daughters (unless it is the mother of a filler-gap construct), the GAP value of the mother is identified with that of the lexical head. As a result, the singleton member of the GAP list of *on* is indirectly passed to the mother *buttered on* and will require the filler NP higher in the tree to be *side*-headed. *On*'s index, *y*, appears as the LOCATION of the *butttered[satisfy]fr* in the FRAMES value of the mother (and head daughter).

The Predicational Head-Complement Construction combines idiom *buttered* and the idiom preposition *on* to license the gap-containing, non-finite, Passive VP *buttered on*. This same construction combines the finite passive auxiliary *is* with *buttered on* to license the gap-containing, finite, passive VP *is buttered on*. The Subject-Predicate construction combines the NP *the bread* with the gap-containing VP *is buttered on* to form the gap-containing clause *the bread is buttered on*. This is summarized in the analysis tree shown in Figure 3. Finally, the Non-Subject Interrogative Filler-Gap Construction allows the filler NP *what side* to combine with the gapped clause to license the non-gapped clause *what side the bread is buttered on*, as shown in Figure 4.

---

[24]In reading the following discussion the reader familiar with neither SBCG nor HPSG may wish to consult Sag 2010a for technical details.

[25]This follows the analysis of Sag 2010b, where each predicate 'amalgamates' the GAP values of its arguments into its own GAP value. For an alternative compatible analysis, see Chaves 2012.

$$
\begin{bmatrix}
\text{FORM } \langle\text{the, bread, is, buttered, on}\rangle \\
\text{SYN} \quad S[\text{GAP } \langle\boxed{1}\, NP_y[\text{LID } \langle\textit{i-side[loc]}(y)\rangle]\rangle] \\
\text{SEM} \quad \begin{bmatrix}\text{FRAMES} \left\langle \begin{array}{l} \boxed{4}\, l_3\text{:}\textit{the}(x, l_2, l_1),\ \boxed{5}l_2\text{:}\textit{i-bread[needs]}(x), \\ \boxed{2}l_1\text{:}\textit{buttered[satisfy]}(s, x, y) \end{array} \right\rangle \end{bmatrix}
\end{bmatrix}
$$

$$
\boxed{3}\begin{bmatrix}
\text{FORM } \langle\text{the, bread}\rangle \\
\text{SYN} \quad NP \\
\text{SEM} \quad \begin{bmatrix}\text{IND} \quad x \\ \text{FRAMES } \langle\boxed{4},\boxed{5}\rangle\end{bmatrix}
\end{bmatrix}
\qquad
\textbf{H:}\begin{bmatrix}
\text{FORM } \langle\text{is, buttered, on}\rangle \\
\text{SYN} \quad VP\begin{bmatrix}\text{GAP } \langle\boxed{1}\rangle \\ \text{VAL } \langle\boxed{3}NP_x\rangle\end{bmatrix} \\
\text{SEM} \quad \begin{bmatrix}\text{FRAMES } \langle\boxed{2}\rangle\end{bmatrix}
\end{bmatrix}
$$

$$
\textbf{H:}\begin{bmatrix}
\text{FORM } \langle\text{is}\rangle \\
\text{SYN} \quad V\begin{bmatrix}\text{GAP } \langle\boxed{1}\rangle \\ \text{VAL } \langle\boxed{3},\boxed{6}\rangle\end{bmatrix} \\
\text{SEM} \quad \begin{bmatrix}\text{FRAMES } \langle\rangle\end{bmatrix}
\end{bmatrix}
\qquad
\boxed{6}\begin{bmatrix}
\text{FORM } \langle\text{buttered, on}\rangle \\
\text{SYN} \quad VP\begin{bmatrix}\text{CAT}\begin{bmatrix}\text{LID } \langle\boxed{2}\rangle \\ \text{VF } \textit{pas}\end{bmatrix} \\ \text{VAL } \langle\boxed{3}\rangle \\ \text{GAP } \langle\boxed{1}\rangle\end{bmatrix} \\
\text{SEM} \quad \begin{bmatrix}\text{FRAMES } \langle\,\boxed{2}\,\rangle\end{bmatrix}
\end{bmatrix}
$$

Figure 3: Analysis of the Clause: *the bread is buttered on*

$$\begin{bmatrix} \text{FORM} & \langle \text{what, side, the, bread, is, buttered, on} \rangle \\ \text{SYN} & \text{S}[\text{GAP } \langle \rangle] \\ \\ \text{SEM} & \begin{bmatrix} \text{FRAMES} & \left\langle \begin{array}{l} \boxed{8}\textit{what}(y,l_7,l_3), \boxed{7}l_7\text{:}\textit{i-side[loc]}(y), \boxed{4}l_3\text{:}\textit{the}(x,l_2,l_1), \\ \boxed{5}l_2\text{:}\textit{i-bread[needs]}(x), \boxed{2}l_1\text{:}\textit{buttered[satisfy]}(s,x,y) \end{array} \right\rangle \end{bmatrix} \end{bmatrix}$$

$$\boxed{1}\begin{bmatrix} \text{FORM} & \langle \text{what, side} \rangle \\ \text{SYN} & \text{NP}[\text{LID } \langle \boxed{7} \rangle] \\ \text{SEM} & \begin{bmatrix} \text{IND} & y \\ \text{FRAMES} & \langle \boxed{8},\boxed{7} \rangle \end{bmatrix} \end{bmatrix} \quad \textbf{H:} \begin{bmatrix} \text{FORM} & \langle \text{the, bread, is, buttered, on} \rangle \\ \text{SYN} & \text{S}[\text{GAP } \langle \boxed{1} \rangle] \\ \text{SEM} & \begin{bmatrix} \text{FRAMES} & \langle \boxed{4}, \boxed{5}, \boxed{2} \rangle \end{bmatrix} \end{bmatrix}$$
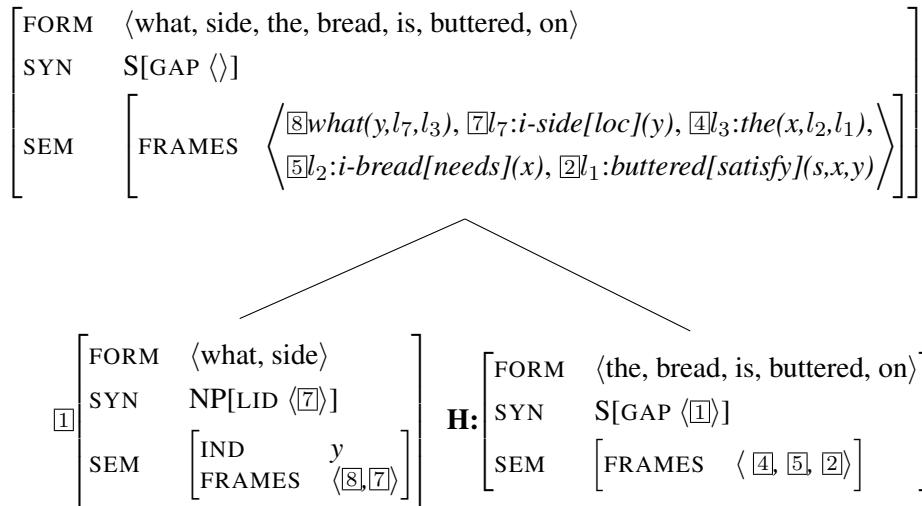
Figure 4: The NonSubject *Wh*-Interrogative Clause: *what side the bread is buttered on*

4.3 What's X Doing Y?

The MRS analysis of *wh*-interrogatives used in the previous section also provides a natural home for a treatment of the 'What's X Doing Y?' (WXDY) Construction, illustrated in (40), whose importance was first discussed by Kay and Fillmore (1999):

(40) a. What were you doing talking to that awful man?

   'Why were you talking to that awful man?'

   b. What is your name doing in my book? (Kay and Fillmore 1999, 3)

   c. 'How come your name is in my book?'

The semantic fact of particular importance in Kay and Fillmore's discussion is the interpretation paraphrasable in terms of *why*, *how come* or *what is the reason that*, as indicated in (40).

   The essential ingredients of WXDY are the following:

(41) a. an interrogative filler *what* participating in a *wh*-interrogative construction,

30

b. a form of the copula governing *doing*,

c. a gap associated with the object of the progressive participle of the verb *do*,

d. a predicative XP following *doing*, forming a constituent with it,

e. the impossibilty of negation, either of *be* or of *do*,

f. a causal interrogative semantics, and

g. a pragmatic attribution of incongruity of the proposition whose cause is being questioned.

These points are illustrated by the following examples:

(42) a. I wonder what the salesman will say this house is doing without a kitchen. (invented example, Kay and Fillmore 1999, ex. (3)c)

b.*What does your name keep doing in my book?

c.*What will your name (be) do in my book?

d. What is he doing? (lacks WXDY semantics)

e.*What weren't you doing (not) talking to my aunt?

f.#What is he doing drunk, which everyone knew he would be?

Example (42a) is of particular importance in showing that the scope of the causal operator is not necessarily the same as the clause following the *what*. That is, though the position of *what* demarcates the top of the interrogative clause, it is the embedded structure *this house is doing without a kitchen* whose causality is to be explained by the salesman. (42a) does not mean 'I wonder why it is that the salesman will say that this house lacks a kitchen'.

WXDY finds a simple analysis within SBCG. Perhaps unsurprisingly, this analysis, like the previous ones, is purely lexical in nature. First, in order to account for the role of *be* in WXDY, we posit a listeme like the following:

(43)
$$
\begin{bmatrix}
\textit{copula-lxm} \\
\\
\text{ARG-ST} \quad \left\langle \text{X}, \begin{bmatrix} \text{VP} \\ \text{LID} \ \langle \textit{i-doing-fr} \rangle \end{bmatrix} \right\rangle
\end{bmatrix}
$$

This listeme, which inherits all of its remaining properties from the Copula Construction (a lexical class construction), selects a subject (X) and a VP complement whose LID is the idiomatic *i-doing-frame*. Like other copula *be*-lexemes, this is an auxiliary verb with subject-raising properties. And because its FRAMES list is empty, it makes no contribution to the semantics.

The lexicon contains only one listeme whose LID is *i-doing-fr*, and hence only one lexeme that gives rise to words that can head the VP complement of the *be* in (43). This listeme, because it includes the specification [VF *prp*], will have only one kind of word realization—a present participle as the one sketched in (44):

(44)

$$
\begin{bmatrix}
\textit{word} \\
\text{FORM} \quad \langle \text{doing} \rangle \\
\text{ARG-ST} \quad \left\langle \boxed{1}, \boxed{2}, \boxed{3}\text{XP}\begin{bmatrix} \text{VAL} & \langle\boxed{1}\rangle \\ \text{LTOP} & l \end{bmatrix} \right\rangle \\
\text{SYN} \quad \begin{bmatrix} \text{CAT} & \begin{bmatrix} \textit{verb} \\ \text{LID} & \langle \textit{i-doing-fr} \rangle \\ \text{VF} & \textit{prp} \end{bmatrix} \\ \text{VAL} & \langle \boxed{1}, \boxed{3} \rangle \\ \text{GAP} & \langle \boxed{2}\text{NP}_x[\text{LID}\langle \textit{what-fr}\rangle] \rangle \end{bmatrix} \\
\text{SEM} \quad \begin{bmatrix} \text{INDEX} & x \\ \text{FRAMES} & \left\langle \begin{bmatrix} \textit{justification-fr} \\ \text{EXPLICANS} & x \\ \text{EXPLICANDUM} & l \end{bmatrix} \right\rangle \end{bmatrix}
\end{bmatrix}
$$

The ARG-ST list of the verb in (44) contains three elements: a subject (□) a direct object (□) and a predicational phrase (□). The direct object, however, is absent from the VAL list and present on the verb's GAP list. This element is specified as [INDEX x] and [LID $\langle$*what-fr*$\rangle$] assuring that the filler daughter will also be so specified. The relevant properties of the word *what* are specified in (45).

$$
(45) \quad
\begin{bmatrix}
\textit{word} \\
\text{FORM} \quad \big\langle \text{what} \big\rangle \\[6pt]
\text{SYN} \quad
\begin{bmatrix}
\text{CAT} \quad
\begin{bmatrix}
\textit{noun} \\
\text{LID} \quad \boxed{1}\big\langle \textit{what-fr} \big\rangle \\
\text{SELECT} \quad \textit{none}
\end{bmatrix} \\[12pt]
\text{WH} \quad \boxed{1} \\
\text{REL} \quad \{\,\}
\end{bmatrix} \\[30pt]
\text{SEM} \quad
\begin{bmatrix}
\text{INDEX} \quad x \\[6pt]
\text{FRAMES} \quad
\left\langle
\begin{bmatrix}
\textit{what-fr} \\
\text{LABEL} \quad \textit{label} \\
\text{BV} \quad x \\
\text{REST} \quad l_1 \\
\text{SCOPE} \quad \textit{lable}
\end{bmatrix},
\begin{bmatrix}
\textit{thing—fr} \\
\text{LABEL} \quad l_1 \\
\text{INST} \quad x
\end{bmatrix}
\right\rangle
\end{bmatrix}
\end{bmatrix}
$$
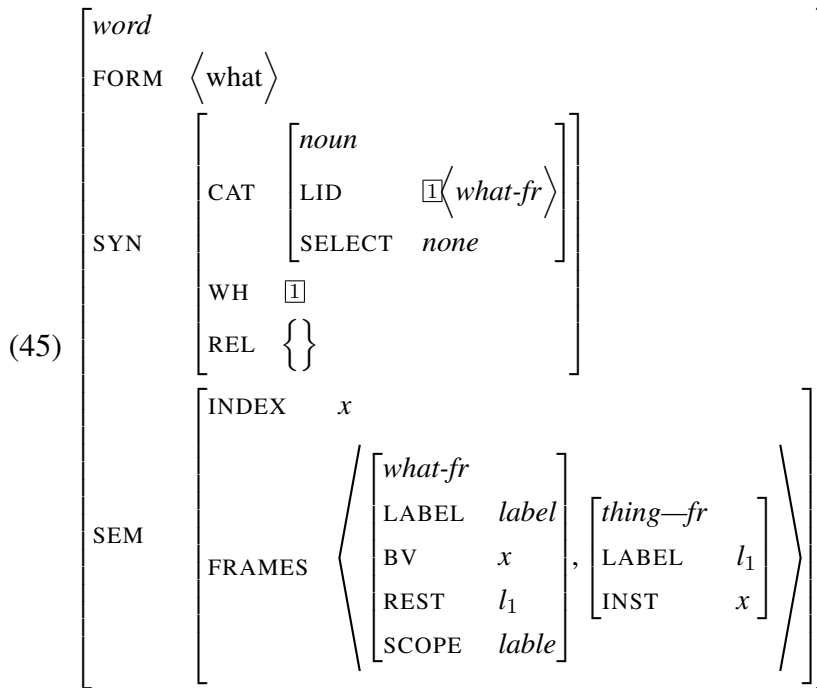
Finally, by couching the semantics in terms of a *justification-frame*, we encode the pragmatic incongruity effect observed by Kay and Fillmore without further stipulation. Note that, aside from register difference, (46a) and (46b) have quite similar perlocutionary effect.

(46) a. What is your name doing in my book? (=40b)

    b. What is the justification of your name being in my book?

In sum, the WXDY Construction is analyzed in terms of a *be* listeme that selects for a complement whose LID is *i-doing-fr*. Because words like the participle in (44) are the only kind that mention *i-doing-fr*, these are the only words available to serve as lexical heads of the VP complement of *be* in WXDY. But since words like (44) have a nonempty GAP list, they must appear at the bottom of a filler-gap dependency. Moreover, since the first member of their GAP list must have the properties of interrogative *what*, that filler-gap dependency must be a *wh*-interrogative clause (main or embedded) whose filler daughter is the word *what*. (44) also links things together to produce a semantics asking about the justification for a certain proposition, where that proposition is constructed from the

subject of *do* (which is also the subject of *be*) and its final complement. A construct illustrating WXDY, a FS of type *ns-wh-int-cl*, is shown in Figure 5. The semantics is indicated in abbreviated notation in (47b):

(47) a. What is Bo doing here?

b. *what(x, l_3, l_1), l_3:thing(x), l_1:justification(x,l_2), l_2:location(Bo,here)*

c. 'What justifies Bo being here?'

5 Locality and Idiomaticity

We have not yet addressed an additional problem that arises in the idiom data. Certain idioms contain obligatory modifiers, as well as obligatory determiners. An example is (48):

(48) bark up the wrong tree.

In this idiom it is possible to insert an optional modifier between *wrong* and *tree*, as shown in (49). The possibility of such modifiers forecloses the possibility of treating the whole idiom as a 'word with spaces'.

(49) a. Barking up the wrong avocado tree in search of authentic Mexican... flavors

b. Barking up the wrong evidence tree...

c. Could we have been barking up the wrong linguistic tree all these years by over-emphasizing the importance of complexity and accuracy?

d. It isn't always possible to avoid wrong turns, but these steps may help keep you from barking up the wrong family tree.

Another example of an idiom that includes obligatory modifiers is *sing(ing) a different tune*, as exemplified in (50).

(50) a. The Sox are singing a different, quieter tune ahead of spring training.

b. However, after spotting a few emerald green dresses underneath the mistletoe last year, I may be singing a different (Christmas) tune...

$$
\textit{wh-int-ns-cl} \quad
\begin{bmatrix}
\text{FORM} \ \langle \text{what, is, Bo, doing, here} \rangle \\[4pt]
\text{SYN} \quad \text{S}
\begin{bmatrix}
\text{CAT} \begin{bmatrix} \text{VF} & \textit{fin} \\ \text{IC} & + \\ \text{INV} & + \end{bmatrix} \\[12pt]
\text{GAP} \ \langle \rangle
\end{bmatrix} \\[30pt]
\text{SEM}
\begin{bmatrix}
\text{LTOP} \quad l_0 \\[6pt]
\text{FRAMES} \ \Bigg\langle
\boxed{1}\begin{bmatrix} \textit{question-fr} \\ \text{LABEL} \ l_0 \\ \text{BV} \quad x \\ \text{REST} \quad l_1 \\ \text{SCOPE} \ l_2 \end{bmatrix},
\boxed{2}\begin{bmatrix} \textit{thing-fr} \\ \text{LABEL} \ l_1 \\ \text{INST} \quad x \end{bmatrix}, \\[40pt]
\boxed{3}\begin{bmatrix} \textit{justification-fr} \\ \text{LABEL} \qquad l_2 \\ \text{EXPLICANS} \quad x \\ \text{EXPLICANDUM} \ l_3 \end{bmatrix},
\boxed{4}\begin{bmatrix} \textit{location-fr} \\ \text{LABEL} \qquad l_3 \\ \text{LOCATUM} \ \textit{Bo} \\ \text{LOCATION} \ \textit{here} \end{bmatrix}
\Bigg\rangle
\end{bmatrix}
\end{bmatrix}
$$

$$
\begin{bmatrix}
\text{FORM} \ \langle \text{what} \rangle \\[4pt]
\text{SYN} \quad \text{NP} \begin{bmatrix} \text{WH} \ \boxed{1} \\ \text{REL} \ \{\} \end{bmatrix} \\[12pt]
\text{SEM} \quad \begin{bmatrix} \text{IND} \qquad x \\ \text{FRAMES} \ \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix}
\end{bmatrix}
\qquad
\begin{bmatrix}
\text{FORM} \ \langle \text{is, Bo, doing, here} \rangle \\[4pt]
\text{SYN} \quad \text{S}
\begin{bmatrix}
\text{CAT} \begin{bmatrix} \text{VF} & \textit{fin} \\ \text{IC} & + \\ \text{INV} & + \end{bmatrix} \\[12pt]
\text{GAP} \ \langle \text{NP}_x \rangle
\end{bmatrix} \\[20pt]
\text{SEM} \quad \begin{bmatrix} \text{FRAMES} \ \langle \boxed{3}, \boxed{4} \rangle \end{bmatrix}
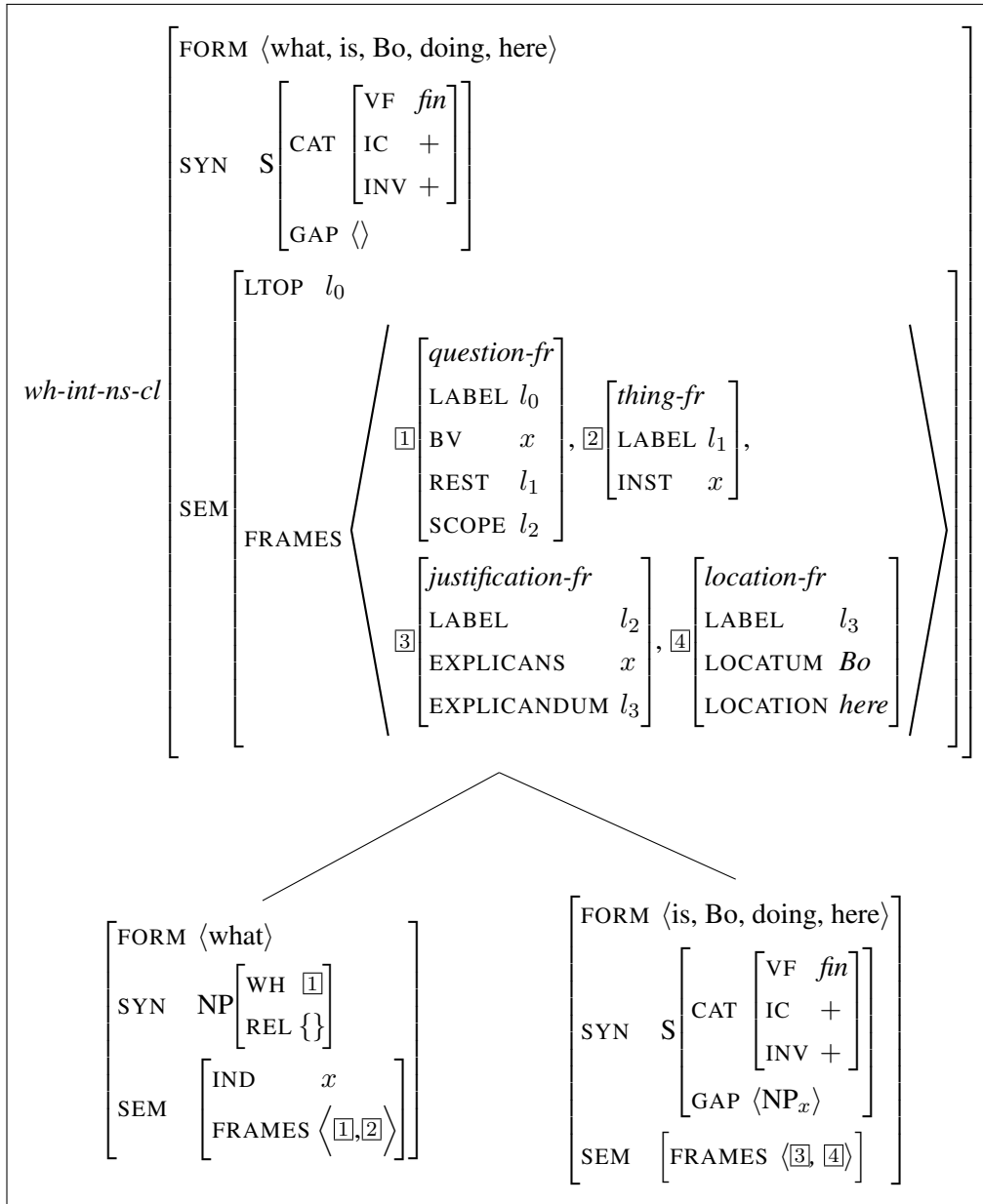\end{bmatrix}
$$

Figure 5: A Construct Illustrating the WXDY Phenomenon

c. ... today's deal from Light Touch Aesthetics for a restorative body wrap will have you singing a different holiday tune.

d. To most Americans, this school marm brandishing a switch seemed to be singing a different, even principled, tune.

The problem that facts like these raises for approaches like the HPSG analysis of Pollard and Sag (1994) or the SBCG treatment sketched in (Sag 2012), is that the Head-Functor analysis of noun phrases, provides no way for a common noun to require a modifier that stands in the functor relation to it. The main predicator, e.g. idiomatic *bark* can require a PP complement headed by idiomatic *up* and idiomatic *up* can require an object headed by idiomatic *tree* but idiomatic *tree* cannot specify that it is modified by idiomatic *wrong*. Indeed, as Richter and Sailer (2009) have shown, this problem of nonlocal dependencies in idiomatic expressions is widespread, crossing clause boundaries in the case of many German idioms that they analyze. In the next section, we show how the same tools we apply here for treating nonlocal idiomatic dependencies in English can provide an account of the German data they discuss.

This problem relates to an issue that we mentioned in passing in section 3, where we proposed the listeme (23) (repeated here) to analyze the semantically empty lexeme *bucket* of *kick the bucket*:

(23) $\begin{bmatrix} cn\text{-}lxm \\ \text{FORM} \quad \langle \text{bucket} \rangle \\ \text{SYN} \quad \begin{bmatrix} \text{CAT} \quad [\text{LID} \ \langle i\text{-}bucket\text{-}fr \rangle] \\ \text{VAL} \quad \langle \rangle \end{bmatrix} \\ \text{SEM} \quad \begin{bmatrix} \text{IND} \qquad none \\ \text{FRAMES} \quad \langle \rangle \end{bmatrix} \end{bmatrix}$

This listeme shows an empty FRAMES list and an LID value of *i-bucket-frame*, unlike the other listemes discussed by Sag (2012), where the FRAMES list and LID value are identified. This raises the question of how to state the general principle and to allow for its exceptions. Given our system of monotonic constraints, this boils down to the question of what class of expressions (i.e. what lexical type) is subject to the identity constraint and what exceptions need to be stated.

The solution to both the problems of modifiers in idioms and obligatory modification in general requires a slight revision in the typology of lexemes. It is convenient to begin by considering the second problem mentioned, obligatory modification. The MRKG feature has til now functioned much like the LID feature, except

that the latter is clearly 'passed up' from head daughter to mother, while MRKG passes up from the non-head daughter to its mother in a head-functor construct. Both features are conceived, however, as providing a partial semantic characterization of the expression they mark. Since the hierarchy representing the range of the LID feature provides a comprehensive semantic taxonomy of lexemes, there is neither formal nor empirical motivation for positing a distinct range of values for the MRKG feature. Any non-maximal classification of a lexeme, such as degree word, *deg*, or definite, *def*, will appear at some level in the frame hierarchy *degree-fr* or *definite-fr*; any maximal classification of a lexeme, such as *i-wrong-fr* will of course also appear there. We simplify the formalism by eliminating a special set of values for the MRKG feature and identifying the MRKG value and LID value for all lexemes.[26]

Thus identifying the MRKG and LID values provides a solution to the obligatory modification problem. Under the proposed change a predicator or functor can specify in its respective ARG-ST or SELECT value a complement or head with a specified MRKG value. This MRKG value of a head can, in the relevant cases, only be acquired from combination with a functor sister bearing that MRKG value in a head-functor construct, and the MRKG value can identify that modifier uniquely if desired. A predicator or potential functor (mainly determiner or modifier) can specify, as loosely or tightly as required, a modifier of the head of the phrase it subcategorizes for or selects. For example, the idiomatic *the* of *up the wrong tree* can specify that the idiom word *tree* it selects has a MRKG value that guarantees that *tree* is modified by *wrong*. Similarly, the idiomatic *up* can subcategorize, not only for a NP whose LID value calls for idiomatic *tree* to be its head, but also that the MRKG value of the phrase it selects indicate the idiomatic *the* which in turn, as we just saw, ensures that the *tree*-headed nominal *it* selects contains the idiomatic modifier *wrong*.

We first consider how the lexemes for *the*, *wrong*, and tree are combined into the phrase *the wrong tree*. (See Figure 6.)[27] The terminal nodes in the analysis tree in Figure 6 display the relevant information about the lexemes *the*,

---

[26]This identification holds *only* for lexemes. We will see how the fact that phrases can, and often do, have different LID and MRKG values provides a strategic advantage.

[27]We assume that *bark*, *up*, *the*, *wrong*, and *tree* are all idiom words; *the* and *wrong*, however, have the same semantics as the corresponding canonical words, while differing in some syntactic properties, specifically LID and MRKG features. We take the semantic breakdown of the idiom to be, roughly, *bark* 'make' *up* (simply raises the semantics of its object) *the* 'the' *wrong* 'wrong' *tree* 'choice'.

$$
\begin{bmatrix}
\text{FORM } \langle \text{the, wrong, tree} \rangle \\[4pt]
\text{SYN} \quad \text{NP}\begin{bmatrix} \text{MRKG } \boxed{8} \\ \text{CAT} \quad [\text{LID } \langle \boxed{1} \rangle] \end{bmatrix} \\[10pt]
\text{SEM} \quad \begin{bmatrix} \text{FRAMES} \left\langle \begin{array}{l} \boxed{8}l_0\text{:}the(x,l_2), \\ \boxed{7}l_2\text{:}i\text{-}wrong(x,l_3), \ \boxed{1}l_3\text{:}i\text{-}tree[choice](x) \end{array} \right\rangle \end{bmatrix}
\end{bmatrix}
$$

$$
\begin{bmatrix}
\text{FORM } \langle \text{the} \rangle \\[4pt]
\text{SYN} \quad \begin{bmatrix} \text{CAT} \begin{bmatrix} \text{LID } \langle \boxed{8} \rangle \\ \text{SEL} \begin{bmatrix} \text{LID} \quad \langle \boxed{1} \rangle \\ \text{MRKG } \boxed{7} \end{bmatrix} \end{bmatrix} \\[10pt] \text{MRKG } \boxed{8} \end{bmatrix} \\[10pt]
\text{SEM} \quad \begin{bmatrix} \text{FRAMES } \langle \boxed{8} \rangle \end{bmatrix}
\end{bmatrix}
\qquad
\begin{bmatrix}
\text{FORM } \langle \text{wrong,tree} \rangle \\[4pt]
\text{SYN} \quad \begin{bmatrix} \text{CAT} \quad [\text{LID } \langle \boxed{1} \rangle] \\ \text{MRKG } \boxed{7} \end{bmatrix} \\[8pt]
\text{SEM} \quad \begin{bmatrix} \text{FRAMES } \langle \boxed{7}, \boxed{1} \rangle \end{bmatrix}
\end{bmatrix}
$$

$$
\begin{bmatrix}
\text{FORM } \langle \text{wrong} \rangle \\[4pt]
\text{SYN} \quad \begin{bmatrix} \text{CAT} \begin{bmatrix} \text{LID } \langle \boxed{7} \rangle \\ \text{SEL} \begin{bmatrix} \text{LID} \quad \langle \boxed{1} \rangle \\ \text{MRKG } \boxed{1} \end{bmatrix} \end{bmatrix} \\[10pt] \text{MRKG } \boxed{7} \end{bmatrix} \\[10pt]
\text{SEM} \quad \begin{bmatrix} \text{FRAMES } \langle \boxed{7} \rangle \end{bmatrix}
\end{bmatrix}
\qquad
\begin{bmatrix}
\text{FORM } \langle \text{tree} \rangle \\[4pt]
\text{SYN} \quad \begin{bmatrix} \text{CAT} \quad [\text{LID } \langle \boxed{1} \rangle] \\ \text{MRKG } \boxed{1} \end{bmatrix} \\[8pt]
\text{SEM} \quad \begin{bmatrix} \text{FRAMES } \langle \boxed{1} \rangle \end{bmatrix}
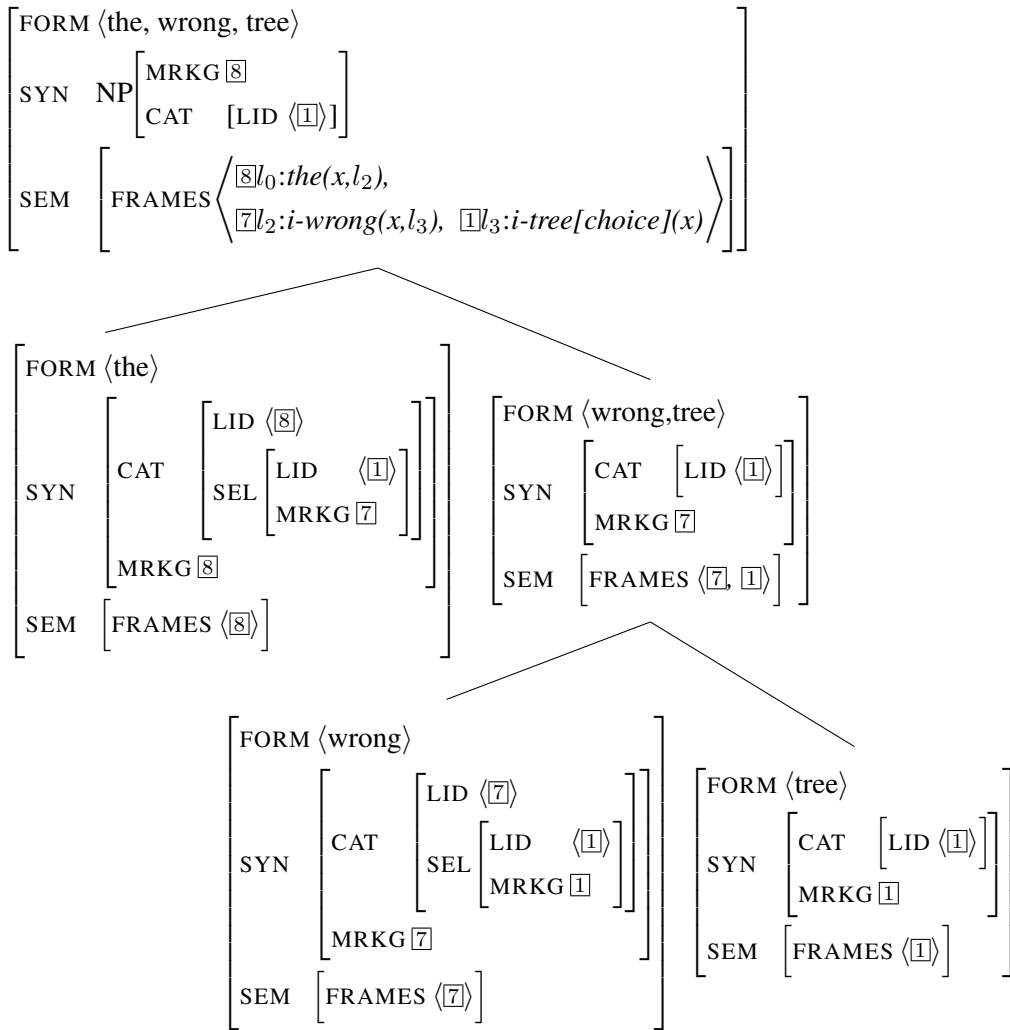\end{bmatrix}
$$

Figure 6: Analysis of *the wrong tree*

wrong, and *tree*.[28] The SELECT value of *wrong* calls for a selectee whose LID and MRKG values both show l_3:*i-tree[choice](x)* The word formed from the idiomatic *tree* lexeme is the only possible expression satisfying this constraint. The mother of the resulting head-functor construct is a phrase specified both as [LID $\langle l_3$:*i-tree[choice](x)* $\rangle$] and [MRKG l_2:*i-wrong(x,l_3)*], the former inherited from the head daughter and the latter from the functor daughter, licensed by the Head-Functor Construction. At the next level, a syntactically idiomatic but semantically canonical lexeme *the* selects a phrase with [LID$\langle l_3$:*i-tree[choice](x)* $\rangle$] and [MRKG *i-wrong-fr*], again licensed by the Head-Functor Construction. The resulting (mother) phrase has [LID $\langle l_3$:*i-tree[choice](x)* $\rangle$] and [MRKG *i-wrong-fr*]. We will see presently that a phrase bearing just these constraints satisfies a valence requirement of idiomatic *up*. We have exemplified how the proposed identification in each listeme of LID and MRKG values in listemes eliminates the otherwise troubling problem of obligatory modification.

Before continuing with the derivation of 'barking up the wrong tree', we propose a further emendation to the framework of Sag 2012. Verbal lexemes are there partially defined by the following lexical class construction:

(51) **Verb Lexeme Construction (↑*lexeme*):**

$$
\textit{verb-lxm} \Rightarrow
\begin{bmatrix}
\text{ARG-ST} & \langle X, \ldots \rangle \\
\text{SYN} & \begin{bmatrix}
\text{CAT} & \begin{bmatrix}
\textit{verb} \\
\text{LID} & L \\
\text{SELECT} & \textit{none} \\
\text{XARG} & X
\end{bmatrix} \\
\text{MRKG} & \textit{unmk}
\end{bmatrix} \\
\text{SEM} & \begin{bmatrix}
\text{LTOP}^{29} & l_{0=_q 1} \\
\text{FRAMES} & L{:}\langle([\text{LABEL} \; l_1])\rangle
\end{bmatrix}
\end{bmatrix}
$$

The constraints on lexemes in general with regard to identity of LID and FRAMES values are not mentioned in Sag 2012, but there are certain classes of lexemes for which this identity cannot hold, as noted above and by Sag in his

---

[28]Note that in the lexemes *the*, *wrong*, and *tree*, the LID and MRKG values are identified, while this is not the case for the phrases *wrong tree* and *the wrong tree*.

[29]The notation [LTOP l_0 ...] indicates that the main frame of the verb will not be outscoped locally by any nonquantifier frame. See Sag 2012, 94 for further discussion. Questions of quantifier scope are not at issue in this paper.

discussion of *kick the bucket*. It is argued that the idiomatic lexeme *bucket* must be formulated with the empty list as the FRAMES value and yet must have an identifying LID distinct from the empty list in order for it to be identifiable by the idiomatic *kick*. Non-predicative ('case-marking') prepositions (including idiomatic ones such as *up* of *up the wrong tree*) must also be formulated with distinct LID values and FRAMES values. In our analysis, which follows a long GPSG/HPSG tradition, the index of the preposition is identified with that of its object. Similarly, infinitive predicative *be* contributes no frames to the meaning of the clauses in which it occurs, and so would also naturally be represented as [FRAMES ⟨⟩].

We slightly amend the lexical branch of the signature hierarchy to (a) unify the LID and MRKG values for all lexemes to account for obligatory modification and (b) properly distinguish common lexemes from empty lexemes: those with empty FRAMES lists and concomitantly distinct FRAMES and LID values.

(52) *lex-sign*
             |
        lexeme

contributing-lxm   *vacuous-lxm*
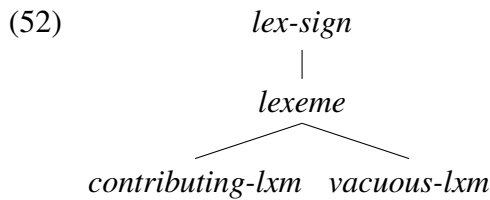
The corresponding type declarations are given in (53):

(53) a. *lex-sign*: [ARG-ST *list(expression)*]

b. *lexeme*:
$$\left[\text{SYN} \begin{bmatrix} \text{CAT} & [\text{LID } \textit{L:list(frame)}] \\ \text{MRKG} & \textit{L} \end{bmatrix} \right]$$

c. *contributing-lxm*:
$$\begin{bmatrix} \text{SYN} & [\text{CAT } [\text{LID } \textit{L}]] \\ \text{SEM} & [\text{FRAMES } \textit{L:nelist(frame)}] \end{bmatrix}$$

d. *vacuous-lxm*:
$$\begin{bmatrix} \text{SYN} & [\text{CAT } [\text{LID } \textit{nelist(frame)}]] \\ \text{SEM} & [\text{FRAMES } \langle\rangle] \end{bmatrix}$$

We note that in (53d) the FRAMES and LID values are not identified, as they are in (53c).

We now complete our analysis of *barking up the wrong tree*. Idiomatic *up* is of type *vacuous-lxm*. This lexeme, since it is specified as [SEM [FRAMES ⟨⟩]], adds no frames to the semantics of the phrase it projects and it concomitantly

displays an LID value, *i-up-fr*, distinct from its FRAMES value, $\langle\rangle$. As mentioned before, idiomatic *up* selects a (noun) phrase with [LID $\langle$*i-tree[choice]-fr*$\rangle$] and [MRKG *the-fr*]. The listeme for idiomatic *up* is given in (54):[30]

(54)
$$
\begin{bmatrix}
\textit{vacuous-lxm \& intran-p-lxm} \\
\text{FORM} \quad \langle\text{up}\rangle \\
\text{SYN} \quad [\text{CAT } [\text{LID } \langle\textit{i-up-fr}\rangle]] \\
\text{ARG-ST} \quad \left\langle \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{CAT} & [\text{LID } \langle\textit{i-tree[choice]-fr}\rangle] \\ \text{MRKG} & \textit{the-fr} \end{bmatrix} \end{bmatrix} \right\rangle
\end{bmatrix}
$$

And this lexeme gives rise to a word that projects a saturated head-complement construct (*sat-hd-comp-cxt*) whose mother's FORM value is $\langle$*up,the,wrong,tree*$\rangle$, as shown in Figure 7.

There is no news in the remainder of our analysis of *bark up the wrong tree*. The idiomatic listeme *bark* is formulated to contribute a frame meaning roughly 'make', to subcategorize for a subject that represents the agent of that frame and a PP, headed by idiomatic *up*, that expresses the incorrect choice made by the agent. The lexeme for idiomatic *bark* is shown in Figure 8.

LID and MRKG are not the only features relevant to the analysis of nonlocal idiomatic dependenciess; XARG has a significant role to play, as well. There are many English idioms that require referential and agreement identity between a genitive within an NP and some other argument of the idiom, or which assign a semantic role to the embedded genitive. Some of these are illustrated in (55)–(56):

(55) a. He$_i$ lost [his$_i$/*her$_j$ marbles].

     b. They$_i$ kept [their$_i$/*our$_j$ cool].

---

[30]Notice that this treatment, like most lexical analyses in the HPSG/SBCG tradition, is highly modular. Further properties of the lexemes instantiating (54) follow from the interaction of other constructions, e.g. (i) and (ii):

(i) Intransitive Preposition Lexeme Construction:

$$
\textit{intran-p-lxm} \Rightarrow \textit{p-lxm} \ \& \ \begin{bmatrix} \text{ARG-ST} & \langle\text{NP}_x\rangle \\ \text{SEM} & [\text{INDEX } x] \end{bmatrix}
$$

(ii) Preposition Lexeme Construction:

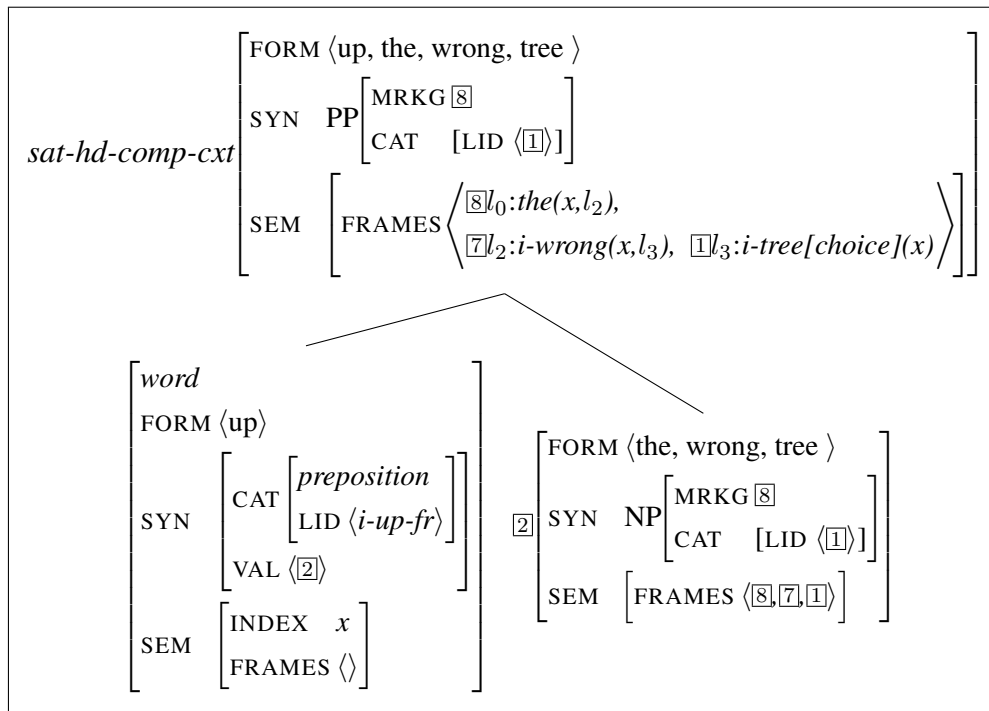$\textit{p-lxm} \Rightarrow [\text{SYN } [\text{CAT } \textit{preposition}]]$

$$
\textit{sat-hd-comp-cxt}
\begin{bmatrix}
\text{FORM } \langle \text{up, the, wrong, tree} \rangle \\[4pt]
\text{SYN} \quad \text{PP}
\begin{bmatrix}
\text{MRKG } \boxed{8} \\
\text{CAT} \quad [\text{LID } \langle \boxed{1} \rangle]
\end{bmatrix} \\[12pt]
\text{SEM}
\begin{bmatrix}
\text{FRAMES}
\left\langle
\begin{array}{l}
\boxed{8} l_0 : \textit{the(x,l}_2\textit{)}, \\
\boxed{7} l_2 : \textit{i-wrong(x,l}_3\textit{)}, \;\; \boxed{1} l_3 : \textit{i-tree[choice](x)}
\end{array}
\right\rangle
\end{bmatrix}
\end{bmatrix}
$$

$$
\begin{bmatrix}
\textit{word} \\
\text{FORM } \langle \text{up} \rangle \\[6pt]
\text{SYN}
\begin{bmatrix}
\text{CAT}
\begin{bmatrix}
\textit{preposition} \\
\text{LID } \langle \textit{i-up-fr} \rangle
\end{bmatrix} \\[8pt]
\text{VAL } \langle \boxed{2} \rangle
\end{bmatrix} \\[10pt]
\text{SEM}
\begin{bmatrix}
\text{INDEX} \quad x \\
\text{FRAMES } \langle \; \rangle
\end{bmatrix}
\end{bmatrix}
\qquad
\boxed{2}
\begin{bmatrix}
\text{FORM } \langle \text{the, wrong, tree} \rangle \\[4pt]
\text{SYN} \quad \text{NP}
\begin{bmatrix}
\text{MRKG } \boxed{8} \\
\text{CAT} \quad [\text{LID } \langle \boxed{1} \rangle]
\end{bmatrix} \\[8pt]
\text{SEM}
\begin{bmatrix}
\text{FRAMES } \langle \boxed{8}, \boxed{7}, \boxed{1} \rangle
\end{bmatrix}
\end{bmatrix}
$$

Figure 7: *up the wrong tree*

$$
\begin{bmatrix}
\textit{prep-intran-v-lxm} \\
\text{FORM} \quad \langle \text{bark} \rangle \\[6pt]
\text{SYN}
\begin{bmatrix}
\text{MRKG} \quad \boxed{1} \\
\text{CAT} \quad [\text{LID } \langle \boxed{1} \rangle] \\
\text{VAL} \quad \langle \text{NP}_y, \text{PP}_x[\text{LID } \langle \textit{i-up-fr} \rangle] \rangle
\end{bmatrix} \\[14pt]
\text{SEM}
\begin{bmatrix}
\text{INDEX} \quad s \\[6pt]
\text{FRAMES}
\left\langle
\boxed{1}
\begin{bmatrix}
\textit{bark[make]-fr} \\
\text{SIT} \quad s \\
\text{AGENT} \quad y \\
\text{THEME} \quad x
\end{bmatrix}
\right\rangle
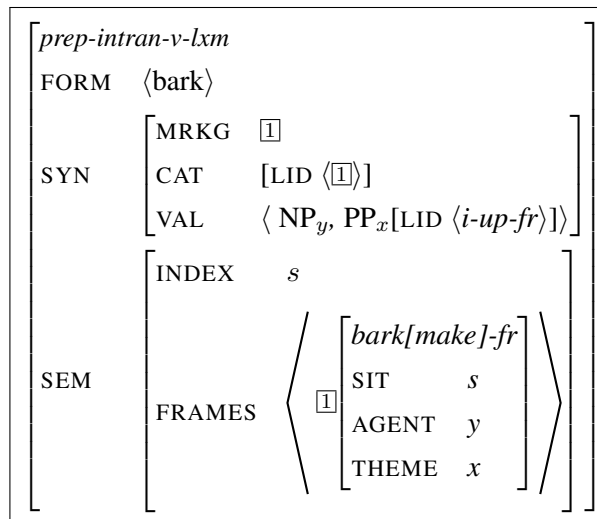\end{bmatrix}
\end{bmatrix}
$$

Figure 8: Idiomatic *bark*

(56) a. Adrenaline$_i$ made [her$_j$ hair] stand on end.

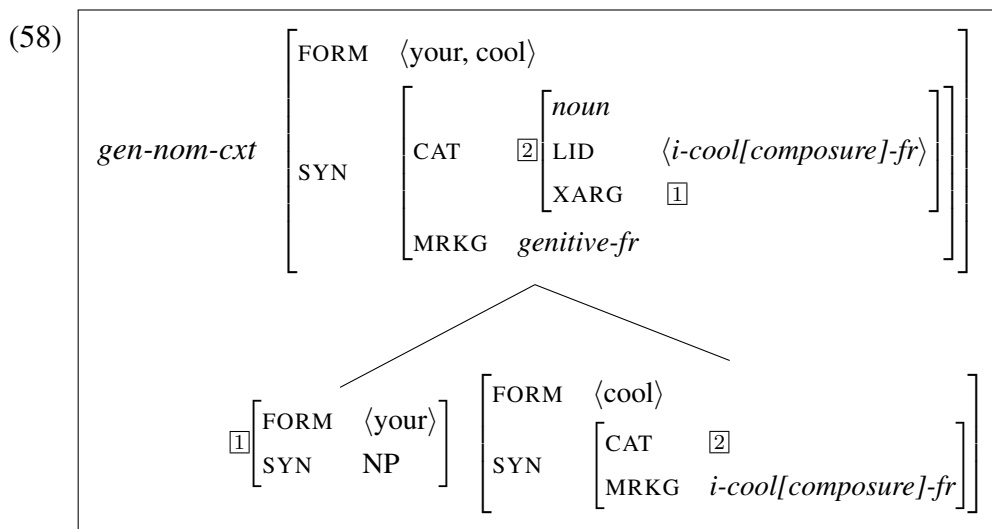   b. This$_i$ tickled [my$_j$ fancy].

As noted in Sag 2012, the presence of a prenominal genitive within an NP is encoded via a nonempty value for the feature XARG. If an object NP includes information about its prenominal genitive in its XARG value, then the listeme of a verb like *lose* (in its 'lose your cool' sense) can be formulated as in (57):[31]

$$
(57) \quad
\begin{bmatrix}
\textit{strans-v-lxm} \\
\text{FORM} \quad \langle \text{lose} \rangle \\
\text{ARG-ST} \quad \left\langle \text{NP}_x \,, \text{NPRV}_y \begin{bmatrix} \text{XARG} & [\textit{pron}]_x \\ \text{LID} & \langle \textit{i-cool[composure]-fr} \rangle \end{bmatrix} \right\rangle \\
\text{SEM} \quad \begin{bmatrix} \text{FRAMES} \quad \left\langle \begin{bmatrix} \textit{losing-fr} \\ \text{AGENT} & x \\ \text{THEME} & y \end{bmatrix} \right\rangle \end{bmatrix}
\end{bmatrix}
$$

This specification requires both that the object NP contain a prenominal pronominal genitive NP and that that pronoun be coindexed with the subject of *lose* (blocking **He lost your cool* and the like).

    We assume that NPs like *your cool* are built via the same Genitive Nominal Construction that is used for such NPs generally. This construction requires that the mother's XARG value be identified with the prenominal genitive NP, as shown in (58):

---

[31]Note that we use '*pron*' as a shorthand for an overt pronominal sign. By contrast, '*pro*' designates a covert pronominal.

(58)

$$
\textit{gen-nom-cxt} \begin{bmatrix} \text{FORM} & \langle \text{your, cool} \rangle \\ \text{SYN} & \begin{bmatrix} \text{CAT} & \boxed{2}\begin{bmatrix} \textit{noun} \\ \text{LID} & \langle \textit{i-cool[composure]-fr} \rangle \\ \text{XARG} & \boxed{1} \end{bmatrix} \\ \text{MRKG} & \textit{genitive-fr} \end{bmatrix} \end{bmatrix}
$$

$$
\boxed{1}\begin{bmatrix} \text{FORM} & \langle \text{your} \rangle \\ \text{SYN} & \text{NP} \end{bmatrix} \quad \begin{bmatrix} \text{FORM} & \langle \text{cool} \rangle \\ \text{SYN} & \begin{bmatrix} \text{CAT} & \boxed{2} \\ \text{MRKG} & \textit{i-cool[composure]-fr} \end{bmatrix} \end{bmatrix}
$$

Thus, because only certain verbs, e.g. *keep*, *lose* and *blow* (in their relevant idiomatic senses) select a direct object whose LID value is ⟨*i-cool-fr*⟩, these are the only lexical elements that can govern NPs headed by *cool* in its relevant idiomatic sense. The genitive within the *cool* NP and the subject of the governor are always coindexed. Various semantic treatments are possible. The lexical entry in (57) assumes that *lose* is dyadic, with the direct object NP forming a second semantic argument (the THEME argument).

The phenomena just discussed are outside the analytic scope of the version of HPSG developed by Pollard and Sag (1994). As argued in Sag (2010b, 2012), these data provide considerable motivation for the analysis of verbal and nominal signs in terms of nonempty XARG specifications. Finally, note that XARG values, unlike VAL lists, do not 'shrink' in a bottom-up progression from head daughter to mother within an analysis tree. That is, no elements are 'cancelled off' an XARG list—the information about the external argument is locally visible at the top of the phrasal domain projected by the lexical head because XARG is a CAT feature and hence is regulated by head feature percolation.

One more idiom further illustrates some of the features of idioms already discussed and also provides a particularly interesting case of a bound genitive pronoun, e.g. *blow one's nose, crane one's neck, get one's comeuppance*. In the idiom illustrated in (59), the verb *keep* involves an AGENT (the keeper of the secret), a THEME (the item kept 'under one's hat', i.e. 'kept secret'), and the *under*-phrase, which specifies where the secret is kept.

(59) a. Maybe Tim kept the best stories under his hat?

b. This has never been shared before, so keep it under your hat.

c. ...so I kept it under my hat until I had enough money to shoot it myself.

We follow the online analysis proposed by The Phrase Finder,[32] where *under one's hat* is decomposed as 'in one's head', i.e. not communicated and hence 'kept secret'.

Notice that there are variant expressions involving *keep* and *under*, e.g. *They kept it under wraps*, *keep it under control,* the idiomaticity of which seems less tightly bound to *keep*. Thus one finds examples like (60a,b):

(60) a. We wonder how long this story was under wraps.

b. Everything is under control.

By contrast, the idiomatic interpretation of *under one's hat* requires the presence of *keep*. The invented examples in (61) have no idiomatic interpretation:[33]

(61) a. This story is under my hat.

b. They have the information under their hat.

Thus our analysis of *keep under one's hat* requires a listeme like (62), where the verb *keep* selects a subject, a direct object, and an *under*-PP whose object must be headed by *hat*.

---

[32]http://www.phrases.org.uk/meanings/under-your-hat.html.
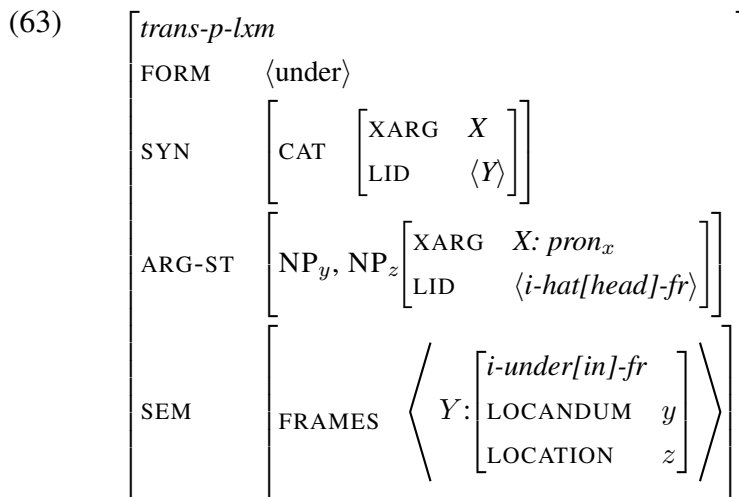
[33]This was apparently not always the case:

(i) The man whose estate lies under his hat need never tremble before the frowns of fortune. [*What I Remember*, by Thomas Trollope (the brother of Anthony), 1887.]

(62)

$$
\begin{bmatrix}
\textit{obj-control-v-lxm} \\
\text{FORM} \quad \langle \text{keep} \rangle \\[4pt]
\text{ARG-ST} \quad \left\langle \text{NP}_x, \text{NP}_y, \text{PP}
\begin{bmatrix}
\text{VAL} & \langle pro_y \rangle \\
\text{LID} & \langle \textit{i-under[in]-fr(y,z)} \rangle \\
\text{XARG} & pron_x \\
\text{LTOP} & l
\end{bmatrix}
\right\rangle \\[4pt]
\text{FRAMES} \quad \left\langle
\begin{bmatrix}
\textit{keep-fr} \\
\text{SIT} & s \\
\text{AGENT} & x \\
\text{THEME} & y \\
\text{STATE} & l
\end{bmatrix}
\right\rangle
\end{bmatrix}
$$
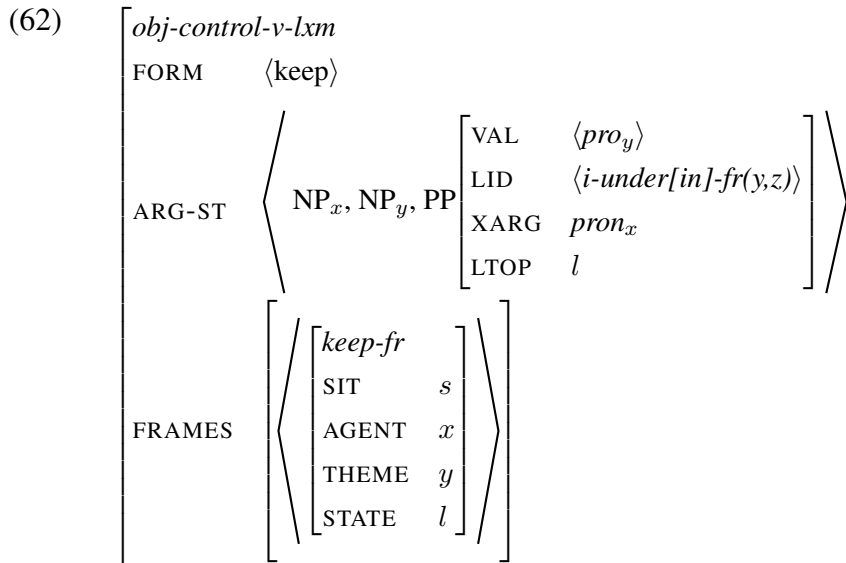
The index y of the first argument of the *i-under[in]-fr* of the PP valent is identified with the index of the object NP$_y$ of keep and the THEME argument of the *keep-fr*. The XARG *pron$_x$* of the PP valent is coindexed with the subject NP$_x$ and its index x is identified with the AGENT argument of the *keep-fr*.

Since *i-under[in]-fr* is an *i-frame*, the idiomatic interpretation will be licensed only when selected by an appropriate governer. In this way, the idiomatic use of the *under*-phrase is restricted to the contexts in which it is governed by *keep*.

This idiomatic lexeme for *under* is sketched in (63):

(63)

$$
\begin{bmatrix}
\textit{trans-p-lxm} \\
\text{FORM} \quad \langle \text{under} \rangle \\[4pt]
\text{SYN} \quad \left[\text{CAT}
\begin{bmatrix}
\text{XARG} & X \\
\text{LID} & \langle Y \rangle
\end{bmatrix}
\right] \\[4pt]
\text{ARG-ST} \quad \left[\text{NP}_y, \text{NP}_z
\begin{bmatrix}
\text{XARG} & X: pron_x \\
\text{LID} & \langle \textit{i-hat[head]-fr} \rangle
\end{bmatrix}
\right] \\[4pt]
\text{SEM} \quad \left[\text{FRAMES} \quad \left\langle Y:
\begin{bmatrix}
\textit{i-under[in]-fr} \\
\text{LOCANDUM} & y \\
\text{LOCATION} & z
\end{bmatrix}
\right\rangle
\right]
\end{bmatrix}
$$

46

This lexeme guarantees that idiomatic *under* governs an *i-hat*-headed NP whose pronominal genitive determiner XARG is identitifed with the XARG of the preposition itself, hence implicitly of the PP it projects, and is coindexed with the subject of *keep*, as sketched in Figure 9.

In Figure 9 the tree structure is determined by the interaction of the valence properties of the individual words and the standard combinatorial constructions of the grammar that build noun phrases, prepositional phrases, and verb phrases. The idiom itself has no phrasal properties. *Your hat* is a canonical noun phrase (determiner phrase, if you prefer); under your hat is a canonical prepositional phrase; and *keep it under you hat* is a canonical verb phrase. The phrase-structural relations observed are determined by the constructions that license such phrases in the grammar generally; no special phrasal machinery is necessary to license the idiom.

Regarding the composition of the semantics in Figure 9, look first at the FRAMES value of the root constituent. The *keep* frame has four arguments: s (a Davidsonian event variable), x, y, and l_2. Note that x is both the index of the subject valent of *keep* and the first argument of the possessor frame ⑤, which identifies the index x of the subject valent and the index x of x's hat (i.e. 'head'). This identification is effected by the joint action of the provision in (62) that identifies the index of the XARG of *keep*'s PP[*under*] argument with that of its subject argument and the property of (63) whereby the tag X identifies the XARG of the preposition *under* (and therefore implicitly the PP *under* projects) with the XARG of its object NP[*hat*]. The y variable is introduced by the *it* constituent, which represents the information kept secret. The l_2 label points to the *i-under[in]-fr*, which says that the secret y is in x's head z. The *the* frame tagged ⑨ is introduced by the *your hat* constituent ④, and represents the definiteness semantics of the possessive determiner. The frame tagged ⑥ assigns the variable z to the index of the *hat* (head) constituent.

6 A Phrasal Alternative[34]

The present SBCG approach to idioms maintains strict locality. It does not permit any listeme or construction to license a dependency more distant than that obtaining within a local tree. It thus conforms to the widespread observation that

---

$$
\begin{bmatrix}
\text{FORM } \langle \text{keep, it, under, your, hat} \rangle \\[4pt]
\text{SYN} \quad \text{VP}
\begin{bmatrix}
\text{MRKG } \boxed{8} \\
\text{CAT} \quad [\text{LID } \langle \boxed{8} \rangle] \\
\text{VAL} \quad \langle \boxed{1}\text{NP}_x \rangle
\end{bmatrix} \\[10pt]
\text{SEM} \quad
\begin{bmatrix}
\text{FRAMES}
\left\langle
\begin{array}{l}
\boxed{8}l_1\text{:}keep(s,x,y,l_2),\ \boxed{7}l_2\text{:}i\text{-}under[in](y,z), \\
\boxed{5}l_4\text{:}poss(x,z)\ ,\ \boxed{9}l_0\text{:}the(z,l_4,l_1)\ ,\ \boxed{6}l_4\text{:}i\text{-}hat[head](z)
\end{array}
\right\rangle
\end{bmatrix}
\end{bmatrix}
$$

$$
\begin{bmatrix}
word \\
\text{FORM } \langle \text{keep} \rangle \\[4pt]
\text{SYN}
\begin{bmatrix}
\text{CAT}
\begin{bmatrix}
verb \\
\text{LID } \langle \boxed{8} \rangle
\end{bmatrix} \\
\text{VAL } \langle \boxed{1},\boxed{2},\boxed{3} \rangle
\end{bmatrix} \\[8pt]
\text{SEM}
\begin{bmatrix}
\text{INDEX} \quad s \\
\text{FRAMES } \langle \boxed{8} \rangle
\end{bmatrix}
\end{bmatrix}
\qquad
\boxed{2}
\begin{bmatrix}
\text{FORM } \langle \text{it} \rangle \\
\text{SYN} \quad \text{NP} \\
\text{SEM } [\text{IND } y]
\end{bmatrix}
\qquad
\boxed{3}
\begin{bmatrix}
\text{FORM } \langle \text{under, your, hat} \rangle \\[4pt]
\text{SYN PP}
\begin{bmatrix}
\text{MRKG } \boxed{7} \\
\text{CAT}
\begin{bmatrix}
\text{LID} \quad \langle \boxed{7} \rangle \\
\text{XARG } \boxed{10}[pron]_x
\end{bmatrix}
\end{bmatrix} \\[8pt]
\text{SEM } \big[\text{FRAMES } \langle \boxed{7},\boxed{5},\boxed{9},\boxed{6} \rangle \big]
\end{bmatrix}
$$

$$
\begin{bmatrix}
word \\
\text{FORM } \langle \text{under} \rangle \\[4pt]
\text{SYN}
\begin{bmatrix}
\text{CAT}
\begin{bmatrix}
preposition \\
\text{LID } \langle \boxed{7} \rangle
\end{bmatrix} \\
\text{VAL } \langle \boxed{4} \rangle
\end{bmatrix} \\[6pt]
\text{SEM } \big[\text{FRAMES } \langle \boxed{7} \rangle \big]
\end{bmatrix}
\qquad
\boxed{4}
\begin{bmatrix}
\text{FORM } \langle \text{your, hat } \rangle \\[4pt]
\text{SYN NP}
\begin{bmatrix}
\text{MRKG } \boxed{8} \\
\text{CAT}
\begin{bmatrix}
\text{LID} \quad \langle \boxed{6} \rangle \\
\text{XARG } \boxed{10}
\end{bmatrix}
\end{bmatrix} \\[8pt]
\text{SEM}
\begin{bmatrix}
\text{INDEX} \quad z \\
\text{FRAMES } \langle \boxed{5},\boxed{9},\boxed{6} \rangle
\end{bmatrix}
\end{bmatrix}
$$
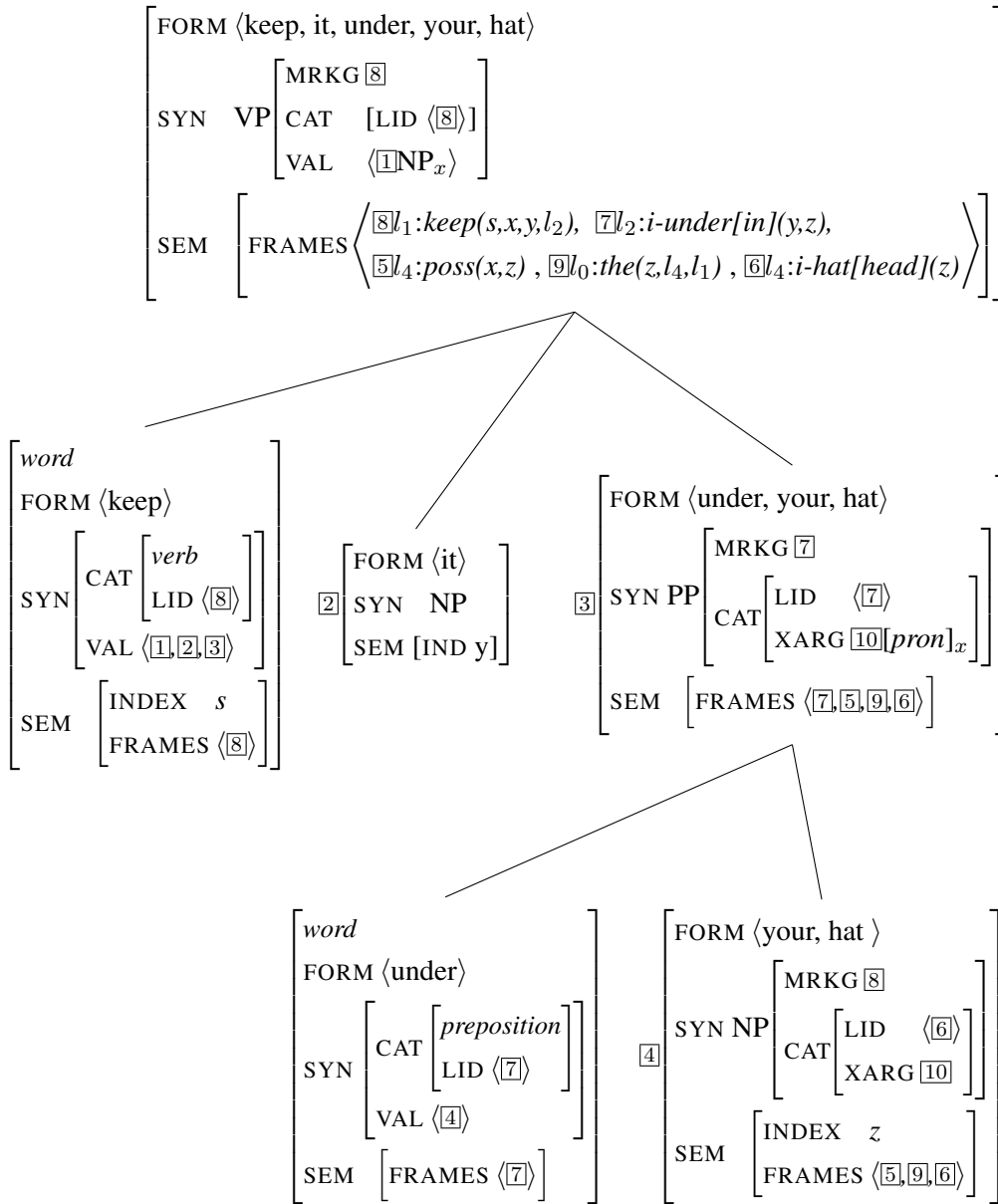
Figure 9: Analysis of: *Keep it Under Your Hat*

48

actual dependencies across the world's languages tend very strongly to be local, apparent exceptions such as extraction presenting in many languages evidence of comprising a series of local dependencies (see e.g., Sag 2007).[35] We conclude that a conservative approach to theory construction seeks a grammatical architecture whose general properties make local dependencies directly expressible and posits additional machinery to express genuinely non-local dependencies only in the special circumstances in which they arise, if any. To adapt an aphorism about simplicity commonly attributed to Einstein, a grammatical theory should be as powerful as it needs to be and no more powerful.

Richter & Sailer (2009 [R&S]), based on earlier work of Sailer (2004) and Soehn (2004), have developed a theory of idioms within constructional HPSG that eschews a lexical and localist approach in favor of a phrasal theory that permits idioms to establish dependencies at an arbitrary depth.[36] Although we find much of value in their treatment of a difficult grammatical question, there is one aspect of their theory that we find unnecessarily powerful. One mechanism by which R&S license such dependencies is by allowing the signs that appear on the DTRS list to contain a DTRS feature. Consequently, a phrasal sign can be defined to specify, for example, that a certain idiom word is a daughter, of a daughter, of a daughter ... of an argument of a specified idiom predicator. The theory of R&S would thus permit the licensing of a *spill-the-beans*-type phrasal idiom with idiomatic listemes *spill* ('divulge') and *beans* ('secrets') that could license sentences such as (64)

(64) a.*Someone spilled that I forgot to remind everyone not to divulge the beans.

b.*Kim spilled that Marion had predicted that Sandy would forget to tell Pat not to divulge the beans.

That is, the R&S theory would permit there to be an idiomatic lexeme *beans* whose entry in the lexicon would specify that *beans* is dominated at an arbitrary

---

[35]To be sure, there exist highly "nonconfigurational" languages, e.g., Warlpiri and other Pama-Nyungan languages, whose loosely constrained word orders require machinery not necessary for English. One example is the HPSG Linearization Theory originally proposed for German by Reape (1994), further developed by several workers, and specifically adapted within a localist and constructional HPSG for Warlpiri by Donohue & Sag (1991; see also the references there).

[36]"Being able to refer to deeply embedded parts of a phrase in [an idiom of a type to be discussed below] is an important ingredient to this theory" (R&S p. 18). It is not clear to us whether R&S intend the suspension of locality to apply only to idioms or to the grammar generally. If the former, they do not specify how the restriction to idioms is to be implemented.

depth by an idiomatic *spill* whose ARG-ST requires only a sentential complement. Of course, the R&S theory would permit the lexicon to refrain from including such items but it would not forbid it. A theory such as the one we have presented, which does not permit the licensing of dependencies of arbitrary depth, captures the generalization that sentences like (64) do not occur.

R&S's abandonment of locality is not casual. They find it to be necessitated by certain facts. The particular data which R&S argue require dependencies of arbitrary depth consist in a number of German idioms in which it appears that coreference is required between a matrix subject and a non-subject constituent of a complement clause. Examples are given in (65), in which the constituent marked X is interpreted as coreferent with the matrix subject.

(65) a. (nicht) wissen, wo X_{dat} der Kopf steht   (R&S 1b)
       not   know   where X_{DO} the head stands
       ('have a lot of stress')

     b. glauben, X_{acc} tritt   ein Pferd   (R&S 1c)
       believe X_{DO} kicks a   horse
       ('be very surprised')

R&S question the adequacy of a theory of the kind proposed here—as briefly sketched in Sag (2012) [cited by R&S in a preprint version dated 2007])—based on the observation that Sag (2012) specifies that the external argument of a predicator is its subject. Consequently non-subject constituents, such as those indicated with X in (65) are not visible to the matrix subject. If SBCG allows only subjects to serve as external arguments and the interpretation of coreference between matrix subject and complement non-subject in (65) is dictated by the grammar rather than the result of a pragmatic process, then the SBCG approach cannot account for the facts of (65). R&S suggest that the SBCG approach might be saved by allowing the ARG-ST to float to the top of a phrase, which would make embedded arguments visible to the matrix subject at the cost of abandoning locality. They continue, however, by presenting facts that they consider to pose a problem for even the anti-locality solution just proposed. Interestingly their evidence for challenging that (to our mind undesirable) solution suggests one of two possible defenses of the strictly localist, SBCG approach. R&S cite the English idiom illustrated in (66). In this case, "X is embedded in a locative modifier. Unless locative modifiers are on the ARG-ST list, the locality assumptions of SBCG do not seem to leave the necessary kind of structure accessible to enforce coreference between X and the matrix subject" (R&S p.19).

(66) look as if butter wouldn't melt [in X's mouth] ('look completely innocent')
(R&S 32)

But is the word *look* part of this idiom? The examples in (67) could be multiplied.

(67) a. This horror novella depicts the fictional life of; [sic] Ethel; a sweet-looking little old lady, who **seems** as if butter wouldn't melt in her mouth...

b. Now Sun is almost 5 and he's just about the cheekiest monkey out there, and Shine is 2 and a half and most of the time **appears** as if butter wouldn't melt in her mouth.

c. ...a brat who pulls all kinds of mischief and then **acts** as if butter wouldn't melt in his mouth.

d. ...where she sits and sulks for several minutes before returning to the bedroom to **pretend** that butter wouldn't melt in her mouth.

e. Look at that innocent expression. ∅ Butter wouldn't melt in his mouth!

f. You lie like ∅ butter wouldn't melt in your mouth, Edward."

g. Justin Bieber—Once upon a time ∅ butter wouldn't melt in little Justin's mouth. Now internationally famous for being a weapons-grade petulant brat...

h. ...he was baited by bullies who made him destroy furniture or school property then pretend he instigated it. ∅ Butter wouldn't melt in their mouth.

In section 4 we cited analogous evidence that—contrary to what one finds frequently claimed—there is no matrix verb in the *side-bread-buttered-on* idiom. Similarly, the idiom illustrated in (66) and (67) is better analyzed as involving the words *butter, wouldn't, melt,* and *mouth,* but not any particular clausal-complement-taking verb, as most clearly demonstrated in (67e,f,g,h). Thomas Wasow has suggested (pc) that idioms of clausal form may carry a pragmatic requirement that one of the constituents therein is contextually bound. In the case of (67h), the contextual antecedent is evidently the bullies. In (67g) there is no controlled pronoun. In (67e,g, and h) there is no higher verb. In (67e) the pragmatically recoverable antecedent is the person the addressee is asked to look at.

51

Another English analogue to the German idioms that appear to require coreference between a matrix verb and a non-subject argument of its complement might be the idiom illustrated in (68).

(68) My friend didn't know what hit him when I poured a bucket of water over his head.

It is common for putatively authoritative sources to include the matrix verb *know* in the idiom. For example, the Cambridge Dictionaries Online gives the idiom as "not know what has hit you." Wictionary lists it as "not know what hit (one)," and the Free Dictionary by Farlex (online) gives it as "not know what hit you." However, the idiom *what hit s.b.* occurs both as a complement to many predicators other than know and freestanding.

(69) a. National Affairs: $\emptyset$ What Hit Him?... What had happened to Harold Stassen in the Nebraska primary?

b. Editorial: Obama still trying to figure out what hit him.

c. "The people are full of anguish, like they don't understand what hit them," Solar TV News reporter David Santos recounted.

d. Ace cueist Pankaj Advani looked uncomfortable in an all-India level final, making faces, raising eyebrows, trying to make out what hit him ...

e. An excellent multiplayer weapon when used with the gauss—most players can't tell what hit them!

f. With the right medical attention, and proper drugging, Lady CBC will not remember what hit her.

g. We were filming a scene where the lndians attack the fort when l suddenly developed a splitting headache. l can't imagine what hit me.

h. Even today, despite numerous works on the crisis—some of them excellent—most Americans remain perplexed by what hit them.

We are thus led to imagine the possibility in the case of (65)b, for example, that the idiom might be restricted to X_{acc} *tritt ein Pferd* plus a pragmatic condition "that it can only be used in contexts describing the mental state of the referent of

the accusative pronoun" (Thomas Wasow pc). We envision the possibility that—for English at least—there may be no idioms that grammatically require coreference of a matrix argument with a non-subject of a complement clause and hence no empirical motivation for a formulation that abandons locality. We tentatively suggest that something of the sort could also be true of German. The first defense of the lexical approach is therefore the suggestion that possibly there are no idioms which require coindexation of a non-subject argument of a complement clause with an argument of a governing verb, the apparent counterexamples containing instead a pragmatic requirement that a pronominal non-subject argument of a complement clause be bound by a contextually given antecedent—which often happens to be an argument (usually the subject) of the matrix predicate.

We have not made a full study of the relevant facts of either German or English (or any other language) and so do not propose this possibility as more than a suggestion for future research. Let us suppose on the contrary that there are in fact idioms which require coreference between a matrix subject (or other argument) and a complement non-subject. We can adjust our theory of idioms to account for this fact (assuming it is a fact) without abandoning locality. Accordingly, we allow a non-subject argument of an idiom predicator—and only of an idiom predicator—to serve as external argument. Idioms are, after all, idiomatic; so it should not occasion great surprise to find that they display a degree of idiosyncratic behavior, especially since the argument made visible from above is usually (if not always) the only "free" argument of the idiom predicator, i.e., the only argument of the idiom predicator not realized by an idiom word or idiom-word-headed phrase. By allowing a non-subject argument of an idiom predicator to serve as external argument we retain locality without sacrificing coverage of inter-clausal coreference.

We stated above that the SBCG grammar signature adopted here distinguishes canonical signs from idiomatic signs. Specifically, the LID value of a canonical sign contains no *i-frames*.

(70) *canonical-sign: sign* & [LID *list(c-frame)*]

An idiomatic sign on the other hand contains at least one *i-frame* in its LID value.

(71) *idiomatic-sign: sign* & [LID $\langle$*i-frame*$\rangle$ $\bigcirc$ *list(frame)*][37]

---

[37]The symbol $\bigcirc$ indicates the *sequence union* (or *shuffle*) operation (Reape 1994). Shuffle constructs a new list from two or more original lists. If $L_1, \ldots L_n$ are lists, then any list containing all and only the members of $L_1, \ldots L_n$ in which the precedence relations in each of $L_1, \ldots L_n$ are conserved is a shuffle of $L_1, \ldots L_n$. (The shuffling together of two or more packets of playing cards provides the image that motivates the name.)

We distinguish *idiom-predicator-verbs* from *canonical-verbs* in order to specify the restricted circumstance in which a non-subject argument can serve as external argument. A canonical verb is defined as (i) having only canonical *signs* on its ARG-ST list and (ii) identifying the XARG with the first item on the ARG-ST list.

(72) **Canonical Verb Construction** ($\uparrow$ *verb-lexeme*)[38]

$$\textit{canonical-v-lxm} \Rightarrow \begin{bmatrix} \text{ARG-ST} & \left\langle \boxed{1}\textit{canonical-sign} \ \oplus \ \textit{list (canonical-sign)} \right\rangle \\ \text{XARG} & \boxed{1} \end{bmatrix}_{39}$$

Idiom-predicator verbs are defined as having ARG-ST lists containing at least one idiomatic sign and allowing any argument to be identified as the XARG. For convenience, we first define an idiomatic-argument-list (*idiom-arg-list*) as a list of signs at least one of which is idiomatic.

(73) *idiom-arg-list:* $\langle \textit{idiom-sign} \rangle \ \bigcirc \ \textit{list(sign)}$

(74) **Idiom Predicator Verb Construction** ($\uparrow$ *verb-lexeme*)

$$\textit{idiom-pred-v-lxm} \Rightarrow \begin{bmatrix} \text{ARG-ST} & \langle \boxed{1}\textit{ sign} \rangle \ \bigcirc \ \textit{idiom-arg-list} \\ \text{XARG} & \boxed{1} \end{bmatrix}$$

R&S analyze example (65)b to illustrate the treatment in their framework of the problematical coreference of the matrix subject with a complement non-subject. The SBCG analysis of this idiom follows. It is convenient to start from the bottom up, with the idiomatic word *Pferd* (lit. 'horse'), which we take to contribute nothing to the meaning of the idiom.[40]

---

[38]This construction will not do for German since German has verbs that lack subjects. "The first nominal argument with structural case is the XARG, if there is any" (Stefan Müller pc).

[39]The symbol $\oplus$ denotes the append operation, which concatenates two lists. Given lists $L_1 = \langle a_1, \cdots, a_n \rangle$ and $L_2 = \langle b_1, \cdots, b_m \rangle$, $L_1 \oplus L_2 = \langle a_1, \cdots, a_n, b_1, \cdots, b_m \rangle$.

[40]We repeat that in claiming the word *Pferd* in this idiom contributes nothing to the meaning of the idiom, we do not suggest that upon hearing or reading the idiom the concept 'horse' is never activated, nor that the metonymy on which the idiom is apparently based is not itself activated for some people. Indeed there is ample evidence from priming studies that exposure to an opaque idiom can prime words related to the meanings of the corresponding non-idiom words; for example *kick the bucket* can prime *pail* (Sprenger et al. 2006, Hillert and Swinney 2001, Colombo 1993, Cacciari and Tabossi 1988, Swinney 1981). (This activation is not observed in all experimental paradigms; see Rommers et al. (2013) for discussion.) On the other hand, the well-known "eggcorn" phenomenon (a *nominal egg* for an *arm and a leg, an utter incomplete fool* for *an utter and complete fool,* etc.) shows that the meanings of fixed phrases, compositional or not, can be learned to some degree independently of the individual words.

$$
\begin{bmatrix}
\textit{cn-lexeme} \\[4pt]
\text{FORM} \quad \left\langle \text{Pferd} \right\rangle \\[6pt]
\text{SYN} \quad \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{LID} \left\langle \textit{i-Pferd-fr} \right\rangle \end{bmatrix} \\ \text{VAL} & \left\langle \right\rangle \end{bmatrix} \\[14pt]
\text{SEM} \quad \begin{bmatrix} \text{INDEX} & \textit{none} \\ \text{FRAMES} & \left\langle \right\rangle \end{bmatrix}
\end{bmatrix}
$$

The remainder of the idiom consists of two listemes, *i-treten* (lit. 'kick') and a listeme which R&S term *surprise-glauben* (hereafter *s-g*), whose paradigm contains forms of both *glauben* and *denken*, the choice of stem determined principally by tense. This listeme "combines with complement clauses that express (negative) surprises, astonishment, or annoyance" (R&S p. 17). *Surprise-glauben* occurs in idioms of this general character, but also with non-idiomatic complements. Examples (75b) are presented by R&S to illustrate both idiomatic (a) and non-idiomatic (b) uses of *s-g*.
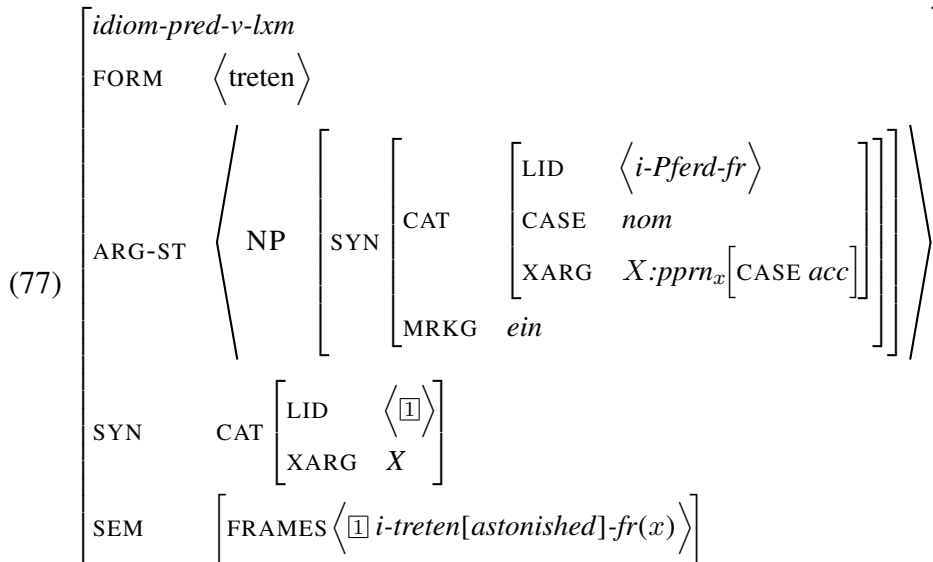
(75) Ich glaub ... [41] (R&S 30)

    a. der hat 'nen Vogel
       he has a   bird
       (... 'he is crazy')

    b. da muss jezt echt alles nochmal neu gemacht werden.
       this must now really all   again   new made   be
       ('... this must all be redone [annoyed]')

R&S insist, however, that *s-g* must be considered part of the *treten-ein-Pferd* idiom. "*Surprise-glauben* is an instance of a (special) attitude predicate that also occurs outside of idioms. For this reason, the matrix predicate need not be restricted to a particular PCl [phraseological clause, e.g., clausal complement of *s-g*]. However, the PCl in Figure 3 [depicting the R&S analysis of the idiom clause: X-acc *treten ein Pferd*] must be collocationally bound to this special matrix predicate..." (p. 17) . In our analysis, the existence of a *surprise-glauben* that takes non-idiom complements along with other *s-g*s that take particular idiom complements is represented in the type hierarchy with a general *s-g* lexeme immediately dominating the *s-g*-lexeme that takes non-idiom complements as well as an additional *s-g* lexeme for each idiomatic complement.
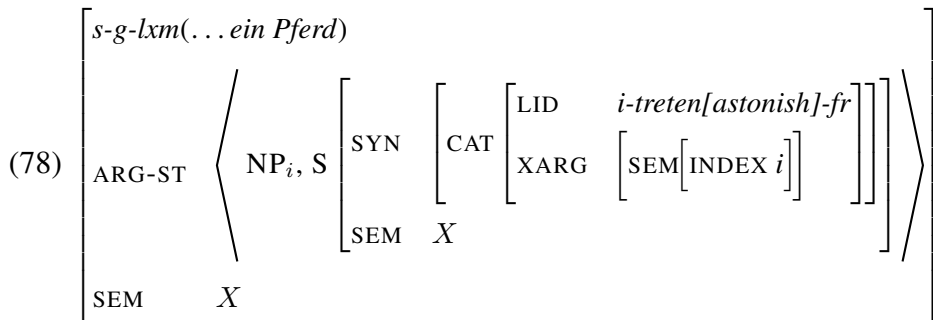
---

[41]We have simplified the example by eliminating some alternative verbs.

(76)

$$s\text{-}g\text{-}lexeme \text{ (general)}$$

$$s\text{-}g\text{-}lxm \text{ (canonical)} \quad s\text{-}g\text{-}lxm \text{ (}\ldots ein\text{-}Pferd\text{)} \quad s\text{-}g\text{-}lxm \text{ (}\ldots \text{ '}nen\text{-}Vogel\text{)} \quad \ldots$$
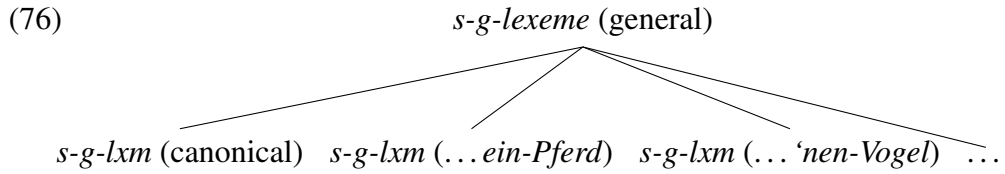
Our analysis attributes the 'surprise' semantics to the idiom-predicator verb *i-treten*, which heads the VP *treten ein Pferd*. (This is analogous to *kick* contributing the meaning 'die' in our analysis of *kick the bucket*.) We specify that *treten-idiom-pred-v-lxm* is dominated in the type hierarchy by *idiom-pred-v-lxm*.

(77)
$$
\begin{bmatrix}
\textit{idiom-pred-v-lxm} \\[4pt]
\text{FORM} \quad \langle \text{treten} \rangle \\[4pt]
\text{ARG-ST} \quad \left\langle \begin{bmatrix} \text{NP} \end{bmatrix} \begin{bmatrix} \text{SYN} \begin{bmatrix} \text{CAT} \begin{bmatrix} \text{LID} & \langle \textit{i-Pferd-fr} \rangle \\ \text{CASE} & \textit{nom} \\ \text{XARG} & \textit{X:pprn}_x \begin{bmatrix} \text{CASE } \textit{acc} \end{bmatrix} \end{bmatrix} \\ \text{MRKG} \quad \textit{ein} \end{bmatrix} \end{bmatrix} \right\rangle \\[4pt]
\text{SYN} \quad \text{CAT} \begin{bmatrix} \text{LID} & \langle \boxed{1} \rangle \\ \text{XARG} & X \end{bmatrix} \\[4pt]
\text{SEM} \quad \begin{bmatrix} \text{FRAMES} \langle \boxed{1} \, \textit{i-treten}[\textit{astonished}]\textit{-fr}(x) \rangle \end{bmatrix}
\end{bmatrix}
$$

The *surprise-glauben* lexeme appropriate to this idiom, *s-g-lxm* (...*ein-Pferd*) subcategorizes for an NP subject. It identifies its subject NP with the XARG of its clausal *i-treten-headed* complement, and identifies its semantics with that of that complement.

(78)
$$
\begin{bmatrix}
\textit{s-g-lxm}(\ldots \textit{ein Pferd}) \\[4pt]
\text{ARG-ST} \quad \left\langle \text{NP}_i, \text{S} \begin{bmatrix} \text{SYN} \begin{bmatrix} \text{CAT} \begin{bmatrix} \text{LID} & \textit{i-treten[astonish]-fr} \\ \text{XARG} & \begin{bmatrix} \text{SEM} \begin{bmatrix} \text{INDEX } i \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{SEM} \quad X \end{bmatrix} \right\rangle \\[4pt]
\text{SEM} \quad X
\end{bmatrix}
$$

To summarize this section: (1) We have suggested that there may exist a pragmatic alternative to R&S's interpretation of certain German idioms as syntactically requiring coreference of a matrix subject and a complement non-subject. We have suggested instead that the coreference observed may represent a pragmatic property (conventional implicature) of these idiomatic constructions. (2) We have shown that if W&S's syntactic interpretation of the observed cases of coreference is correct, that with a small alteration the SBCG lexical and localist approach can account for the data.

Although we disagree with R&S with respect to the matters focused on most closely in this section, we are entirely in sympathy with their (otherwise) successful effort to incorporate into an explicit theory of grammar a theory of idioms that takes account of the extensive and complex facts regarding partial productivity and partial compositionality that idioms present.

7 Idiom Words Governed by Non-Idiom Predicators

Nunberg et al. (1994) cite McCawley's (1981) as pointing out that sentences like (79) and (80) jointly create a problem for the claim that the existence of idioms provides an argument for the existence of transformations. We are not concerned here with arguments for or against transformations but rather with whether examples like these pose a problem for the present lexical approach to idioms.

 (79) Pat pulled [the strings [that got Chris the job]]. (=NSW (33)a)

 (80) [The strings [that Pat pulled]] got Chris the job. (=NSW (33)b)

Example (79) does not cause a problem for the approach to idioms advanced here, because the idiom word *strings* is not governed by a non-idiom predicator.[42]
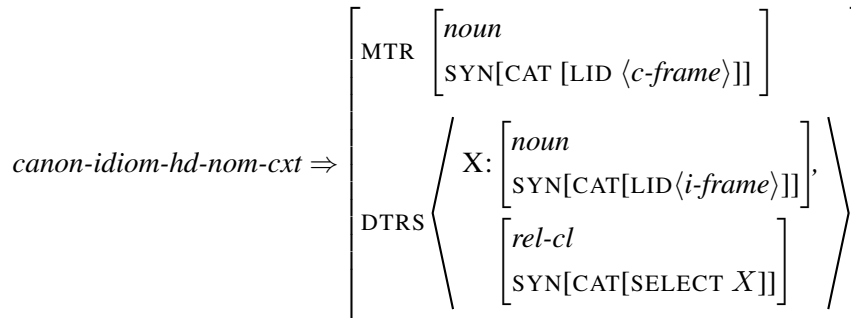
Example (80), however, does present a problem for the current analysis because in this case the nominal phrase headed by *i-strings* is an argument of the canonical predicator got. Speaking informally, *pulled* in the relative clause seems

---

[42]Although (79) and (80) are invented examples, attested ones are easy to find. Compare (i) to (79) and (ii) to (80).

 (i) ... the "big city" bankers who, many people were convinced, pulled the strings that manipulated the political system as well as the economy.

 (ii) ... the daughter was not aware of the strings that were being pulled for her.

somehow to license the token of *i-strings* that is governed in the main clause by *got*. The construction in (81) licenses nominal phrases which, although headed by idiom words, can nonetheless serve as arguments of canonical predicators, for example, *the strings that Pat pulled* in (80).[43]

(81) **Canonical Idiomatically-Headed Nominal Construction (↑*noun-lxm*)**

$$
\textit{canon-idiom-hd-nom-cxt} \Rightarrow
\begin{bmatrix}
\text{MTR} & \begin{bmatrix} \textit{noun} \\ \text{SYN[CAT [LID } \langle \textit{c-frame}\rangle \text{]]} \end{bmatrix} \\
\text{DTRS} & \left\langle \text{X:} \begin{bmatrix} \textit{noun} \\ \text{SYN[CAT[LID}\langle \textit{i-frame}\rangle\text{]]} \end{bmatrix}, \begin{bmatrix} \textit{rel-cl} \\ \text{SYN[CAT[SELECT } X \text{]]} \end{bmatrix} \right\rangle
\end{bmatrix}
$$

The construction in (81) simply codes the fact that a noun or nominal phrase with [LID *i-frame*] can, when modified by a relative clause, head a nominal phrase with [LID *c-frame*] and thus serve as an argument for a canonical predicator, as in (80). If this isolated fact can be found in future research to follow from something more general, that development will of course be welcome. Meanwhile, a constructional approach such as SBCG, while always seeking the smallest number of the most general constructions the facts permit, provides a theoretical vehicle to account for isolated facts of this kind without departing from explicit representation.

8 Conclusion

We have outlined a lexical theory of phrasal idioms. A pervasive property of flexible idioms is that the words that appear in citation forms unmodified are often modified in actual use. This modification is theoretically crucial because it shows that the modified words are not merely words phonologically but also semantically and syntactically. In the following examples, we have confined ourselves to internal modification (ignoring external modifiers like *proverbial, metaphorical, literal, figurative,* etc. and epithets like *danged* or *confounded*.) In the (82) examples idiomatic *cat* is modified; in the (83) examples idiomatic *bag* is modified. We present what may seem like an excessive number of (attested) examples to illustrate the point that ordinary modification of idiom words is commonplace. Modi-

---

[43] And *the strings that were being pulled for her* in (ii) of the preceding note.

fiability of idiom words demonstrates that they are real words with real meanings, which have to be combined to produce meaningful phrases and sentences by the same grammatical machinery that combines the meanings of ordinary words.

(82)   a. I'll bet there might even be more than a few wealthy Palm Beachers who also have not let the **financial** cat out of the bag to everyone in their family.

   b. Now, however, it has to be fleshed out, and in that respect you have not let the **whole** cat out of the bag.

   c. Now, however, it has to be fleshed out, and in that respect you have not let the **whole** cat out of the bag.

   d. We're one day closer to Capcom Cup and it's definitely time to let the **final** cat out of the bag with the Super Street Fighter 4 AE brackets.

   e. Heather said that Blossom almost let the **entire** cat out of the bag today. . . read on if you want a glimpse of what is coming.

   f. Hamid Karzai has let the **Pentagon's** cat out of the bag—to the displeasure of the Obama Administration.

   g. So, without further ado, it brings me great pleasure to finally let the (**first**) cat out of the bag and announce today, the Fritz & Fräulein Collection will be featured and sold at Tommy Hilfiger's beautiful new flagship store,. . .

   h. "Might as well let the **other** cat out of the bag," he said: . . .

   i. Solo let **the Olympic's** cat out of the bag when she narrated her experience to the press.

   j. Don't let the **last** cat out of the bag.

   k. Now, however, it has to be fleshed out, and in that respect you have not let the **whole** cat out of the bag.

   l. We're one day closer to Capcom Cup and it's definitely time to let the **final** cat out of the bag with the Super Street Fighter 4 AE brackets.

   m. Well the liberal Austin bloggers let the **political** cat out of the bag.

n. Georgia, always the truth bringer of the family, let the **cinematic** cat out of the bag, when she allowed that the picture was not such an ordeal as it would have been had their lives been filmed for the entire year they spent on the island. . . .

(83) a. Time to let the cat out of the **Instagram** bag. . . I'm pregnant!

b. Who let the cat out of the **Bloomingdales** bag?

c. Kandi Burruss may have let the cat out of the **Bravo** bag last night during "Watch What Happens Live."

d. Someone let the cat out of the **Duluth Pack** bag a week early.

e. . . . a fetching real estate purveyor down in Miami Beach who kindly let the cat out of the **celebrity real estate** bag about six-time Grammy-winning singer/songwriter Billy Joel. . .

f. So, Mary Burke has finally let the cat out of the **cellophane** bag by formally announcing her candidacy for governor.

g. . . . and this week the folks at Zillow let the cat out of the **real estate** bag . . .

h. Yup, I let the cat out of the **family** bag. Probably the whole reason I am straightforward. . .

i. Now that director Wes Ball has let the cat out of the **concept art** bag,

j. Perhaps sensing that he'd let the cat out of the **plot** bag a little early, King then told Cronenberg and the audience that he wasn't completely committed to the new novel . . .

k. Good morning everyone! Time to let the cat out of the **internet** bag.

l. A little less surprising now you've let the cat out of the **fur-trim, diamante embellished clutch** bag, but we suppose it's the thought that counts.

m. But a guy in Philadelphia (not Mississippi, Pennsylvania this time) let the cat out of the **voter suppression** bag:. . .

n. But some believe it is because some major problems are on the way and NASA does not want other Scientists to let the cat out of the **cosmic** bag to [sic] soon.

The modifiability of idiom words argues for a lexical theory of idioms. Idiom words are real words; they share the morphological and phonological properties of their literal counterparts through multiple inheritance and contribute their own idiomatic meanings. Often these meanings have been motivated historically by metaphors or metonymies of varying degrees of currency. Idiom words are combined with ordinary words and phrases into the phrases and sentences of the grammar by the familiar combinatorial constructions; no special combinatorial machinery is required for idioms. Once idiom words are defined and idiom predicators appropriately distinguished from canonical predicators, nothing further is needed. The ordinary combinatorial constructions that are required for the regular grammar suffice. We suggest that in the past insufficient attention has been paid to the description of idioms within explicit theories of grammar and have attempted to do so here within the approach of Sign-Based Construction Grammar. Our theory of idioms is lexicalist because we find that syntactically flexible idioms are semantically compositional: their meanings are put together from the meanings of special idiom words according to the same processes—combinatorial constructions—that put together the meanings of ordinary phrases and sentences. A welcome result of the commitment to a lexical approach is that little has had to be added to existing theory to account for the data of idioms.

References

Allegranza, Valerio. 2007. *The Signs of Determination. Constraint-Based Modelling Across Languages*, vol. 16 of *SABEST: Saarbrücker Beiträge zur Sprach- und Translationswissenschaft*. Frankfurt am Main: Peter Lang.

Bauer, Laurie. 1983. *English Word-Formation*. Cambridge: Cambridge University Press.

Boas, Hans C. and Ivan A. Sag, eds. 2012. *Sign-Based Construction Grammar*. Stanford, CA: CSLI Publications.

Cacciari, C. and P. Tabossi. 1988. The comprehension of idioms. *Journal of Memory and Language* 27:668–683.

Chaves, Rui Pedro. 2012. On the grammar of extraction and coordination. *Natural Language and Linguistic Theory* 30(2):465–512.

Colombo, L. 1993. The comprehension of ambiguous idioms in context. In C. Cacciari and P. Tabossi, eds., *Idioms: Processing, structure and interpretation*, pages 163–189. Hillsdale, NJ: Erlbaum.

Copestake, Ann, Dan Flickinger, Carl J. Pollard, and Ivan A. Sag. 2005. Minimal recursion semantics: an introduction. *Research on Language and Computation* 3(4):281–332.

Donohue, C. and Ivan A. Sag. 1991. Domains in Warlpiri. Paper presented at HPSG-99 – University of Edinburgh. Non-final draft available at http://www.cs.toronto.edu/ gpenn/csc2517/donohue-sag99.pdf.

Fillmore, Charles J. 1982. Frame semantics. In *In Linguistics in the Morning Calm*, pages 111–137. Seoul: Hanshin Publishing Co.

Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quaderni di Semantica* 6:222–254.

Fillmore, Charles J. 1986. Pragmatically controlled zero anaphora. In *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society*, pages 95–107. BLS.

Fillmore, Charles J. and Colin Baker. 2010. A frames approach to semantic analysis. In B. Heine and H. Narrog, eds., *The Oxford Handbook of Linguistic Analysis*, pages 313–340. Oxford: Oxford University Press.

Fillmore, Charles J., Paul Kay, and Mary C. O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 64:501–538.

Hillert, D. and D. A. Swinney. 2001. The processing of fixed expressions during sentence comprehension. In *Conceptual structure, discourse, and language*, pages 107–121. Stanford, CA: CSLI Publications.

Kay, Paul and Charles Fillmore. 1999. Grammatical constructions and linguistic generalizations: The *what's x doing y?* construction. *Language* 75.1:1–33.

Kay, Paul and Ivan A. Sag. 2012. Discontinuous dependencies and complex determiners. In H. C. Boas and I. A. Sag, eds., *Sign-Based Construction Grammar*, pages 211–238. CSLI Publications.

McCawley, James D. 1981. The syntax and semantics of Enlish relative clauses. *Lingua* 53(2):99–149.

Nunberg, Geoffrey, Ivan A. Sag, and Thomas Wasow. 1994. Idioms. *Language* 70(3):491–538.

Pollard, Carl J. and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Reape, Mike. 1994. Domain union and word order variation in German. In J. Nerbonne, K. Netter, and C. J. Pollard, eds., *German in Head-Driven Phrase Struc-*

ture Grammar, no. 46 in CSLI Lecture Notes, pages 151–197. Stanford University: CSLI Publications.

Richter, Frank and Manfred Sailer. 2009. Phraseological clauses in constructional hpsg. In *Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar, University of Göttingen, Germany*, pages 297–317. Stanford: CSLI Publications.

Rommers, J., T. Dijkstra, and M. C. M. Bastiaansen. 2013. Context-dependent semantic processing in the human brain: Evidence from idiom comprehension. *Journal of Cognitive Neuroscience* 25:762–776.

Sag, Ivan A. 2010a. English filler-gap constructions. *Language* 86:486–545.

Sag, Ivan A. 2010b. Feature geometry and predictions of locality. In G. Corbett and A. Kibort, eds., *Features: Perspectives on a Key Notion in Linguistics*, pages 236–271. Oxford: Clarendon Press.

Sag, Ivan A. 2012. Sign-based construction grammar: An informal synopsis. In H. C. Boas and I. A. Sag, eds., *Sign-Based Construction Grammar*, pages 69–202. Stanford, CA: CSLI Publications.

Sag, Ivan A., Timothy Baldwin, Francis Bond, Ann Copestake, and Daniel P. Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. In *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2002)*, pages 1–15. Mexico City, Mexico.

Sag, Ivan A., Thomas Wasow, and Emily M. Bender. 2003. *Syntactic Theory: A Formal Introduction*. Stanford: CSLI Publications, 2nd edn.

Sailer, Manfred. 2004. Propositional relative clauses in german. In S. Müller, ed., *Proceedings of the HPSG-2004 Conference, Center for Computational Linguistics, Katholieke Universiteit Leuven*, pages 223–243. Stanford: CSLI Publications.

Soehn, Jan-Philipp. 2004. License to COLL. In S. Müller, ed., *Proceedings of the HPSG-2004 Conference, Center for Computational Linguistics, Katholieke Universiteit Leuven*, pages 261–273. Stanford: CSLI Publications.

Sprenger, S. A., W. J. M. Levelt, and G. Kempen. 2006. Lexical access during the production of idiomatic phrases. *Journal of Memory and Language* 54:161–184.

Swinney, D. A. 1981. Lexical processing during sentence comprehension: Effect of higher order constraints and implications for representation. In T. Meyers, J. Laver, and J. Anderson, eds., *The cognitive representation of speech*, Advances in Psychology Series, pages 201–209. Amsterdam: North-Holland.

Van Eynde, Frank. 2006. NP-internal agreement and the structure of the noun

phrase. *Journal of Linguistics* 42(1):139–186.

Van Eynde, Frank. 2007. The big mess construction. In S. Müller, ed., *The Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar, Stanford University*, pages 415–433. Stanford: CSLI Publications.

Wasow, Thomas, Ivan A. Sag, and Geoffrey Nunberg. 1984. Idioms: An interim report. In S. Hattori and K. Inoue, eds., *Proceedings of the 13th International Congress of Linguistics (Tokyo)*, pages 102–115. The Hague: CIPL.